

1987

Image analysis, modeling, enhancement, restoration, feature extraction and their applications in nondestructive evaluation and radio astronomy

Yi Zheng
Iowa State University

Follow this and additional works at: <https://lib.dr.iastate.edu/rtd>



Part of the [Astrophysics and Astronomy Commons](#), and the [Electrical and Electronics Commons](#)

Recommended Citation

Zheng, Yi, "Image analysis, modeling, enhancement, restoration, feature extraction and their applications in nondestructive evaluation and radio astronomy" (1987). *Retrospective Theses and Dissertations*. 9321.
<https://lib.dr.iastate.edu/rtd/9321>

This Dissertation is brought to you for free and open access by the Iowa State University Capstones, Theses and Dissertations at Iowa State University Digital Repository. It has been accepted for inclusion in Retrospective Theses and Dissertations by an authorized administrator of Iowa State University Digital Repository. For more information, please contact digirep@iastate.edu.

INFORMATION TO USERS

The most advanced technology has been used to photograph and reproduce this manuscript from the microfilm master. UMI films the original text directly from the copy submitted. Thus, some dissertation copies are in typewriter face, while others may be from a computer printer.

In the unlikely event that the author did not send UMI a complete manuscript and there are missing pages, these will be noted. Also, if unauthorized copyrighted material had to be removed, a note will indicate the deletion.

Oversize materials (e.g., maps, drawings, charts) are reproduced by sectioning the original, beginning at the upper left-hand corner and continuing from left to right in equal sections with small overlaps. Each oversize page is available as one exposure on a standard 35 mm slide or as a 17" × 23" black and white photographic print for an additional charge.

Photographs included in the original manuscript have been reproduced xerographically in this copy. 35 mm slides or 6" × 9" black and white photographic prints are available for any photographs or illustrations appearing in this copy for an additional charge. Contact UMI directly to order.



Accessing the World's Information since 1938

300 North Zeeb Road, Ann Arbor, MI 48106-1346 USA

Order Number 8805156

**Image analysis, modeling, enhancement, restoration, feature
extraction and their applications in nondestructive evaluation
and radio astronomy**

Zheng, Yi, Ph.D.

Iowa State University, 1987

U·M·I

**300 N. Zeeb Rd.
Ann Arbor, MI 48106**

PLEASE NOTE:

In all cases this material has been filmed in the best possible way from the available copy. Problems encountered with this document have been identified here with a check mark ☒.

1. Glossy photographs or pages ☒
2. Colored illustrations, paper or print ☐
3. Photographs with dark background ☒
4. Illustrations are poor copy ☐
5. Pages with black marks, not original copy ☒
6. Print shows through as there is text on both sides of page ☐
7. Indistinct, broken or small print on several pages ☒
8. Print exceeds margin requirements ☐
9. Tightly bound copy with print lost in spine ☐
10. Computer printout pages with indistinct print ☐
11. Page(s) _____ lacking when material received, and not available from school or author.
12. Page(s) _____ seem to be missing in numbering only as text follows.
13. Two pages numbered _____. Text follows.
14. Curling and wrinkled pages ☐
15. Dissertation contains pages with print at a slant, filmed as received ☐
16. Other _____

U·M·I

Image analysis, modeling, enhancement, restoration, feature
extraction and their applications in nondestructive evaluation and
radio astronomy

by

Yi Zheng

A Dissertation Submitted to the
Graduate Faculty in Partial Fulfillment of the
Requirements for the Degree of
DOCTOR OF PHILOSOPHY

Department: Electrical Engineering and Computer Engineering
Major: Electrical Engineering
(Communications and Signal Processing)

Approved:

Signature was redacted for privacy.

In Charge of Major Work

Signature was redacted for privacy.

For the Major Department

Signature was redacted for privacy.

For the Graduate College

Iowa State University
Ames, Iowa

1987

TABLE OF CONTENTS

	PAGE
GENERAL INTRODUCTION	1
PART I. MODELING VERY LARGE ARRAY PHASE DATA BY THE BOX- JENKINS METHOD	6
ABSTRACT	7
INTRODUCTION	8
DATA CHARACTERISTICS	14
TIME-SERIES MODELS OF THE PHASE	28
DISCUSSION	48
ACKNOWLEDGMENTS	52
REFERENCES	53
PART II. IMAGE ANALYSIS, FEATURE EXTRACTION, AND VARIOUS APPLIED ENHANCEMENT METHODS FOR NDE X-RAY IMAGES	56
ABSTRACT	57
INTRODUCTION	58
X-RAY RADIOGRAPHIC IMAGE ANALYSIS	60
Image Contrast	60
Unsharpness	62
Film Graininess	63
ENHANCEMENT METHODS	67
Trend Removal	67
Dynamic Range Expansion	68
Histogram Equalization	69
Median Filter	69
Sigma Filter	70
Adaptive Smoothing Filter	71
Kalman Filter for Image Restoration	72
IMAGE FEATURE EXTRACTION	74
Flaw Spatial Activity Detection	74
Flaw Detection by Image Segmentation	77
ACKNOWLEDGMENTS	79
REFERENCES	80

PART III. APPLICATION OF ADAPTIVE REGIONAL KALMAN FILTERING TO X-RAY IMAGES IN NDE	84
ABSTRACT	85
INTRODUCTION	86
SEGMENTATION	88
MODELING	91
KALMAN FILTERING	93
RESULTS	95
ACKNOWLEDGMENT	97
REFERENCES	98
PART IV. AUTOMATIC IMAGE SEGMENTATION, MODELING AND RESTORATION WITH A RULE-BASED EXPERT SYSTEM	99
ABSTRACT	100
INTRODUCTION	101
ALGORITHMS	103
Image Segmentation	103
Noise Measurement and Image Modeling	104
Adaptive Kalman Filtering	106
RULE-BASED EXPERT SYSTEM	108
EXPERIMENTAL RESULTS	113
ACKNOWLEDGMENTS	118
REFERENCES	119
PART V. NDE X-RAY IMAGE MODELING AND ADAPTIVE FILTERING CONSIDERING CORRELATED NOISE	121
ABSTRACT	122
INTRODUCTION	123
X-RAY RADIOGRAPHIC IMAGE MODELING	125
IMAGE ESTIMATION BY ADAPTIVE KALMAN FILTERING	132

ENHANCING INDUSTRIAL RADIOGRAPHIC X-RAY IMAGES	136
ACKNOWLEDGMENT	139
REFERENCES	140
PART VI. KALMAN FILTERING VLA PHASE DATA WITH A SUPERCOMPUTER	143
PROBLEM DEFINITION	144
STANDARD CORRECTION TECHNIQUES	145
ARIMA MODELING AND KALMAN FILTERING METHOD	146
COMPUTATIONAL TIME PROBLEM	148
OPTIMIZATION AND VECTORIZATION OF THE KALMAN FILTERING ROUTINE ON THE BCS CRAY X-MP/24 COMPUTER	150
CONCLUSION	154
REFERENCES	155
PART VII. DEBLURRING NDE X-RAY IMAGES AND INFRARED IMAGES BY A MAXIMUM ENTROPY DECONVOLUTION METHOD	156
ABSTRACT	157
INTRODUCTION	158
THE MAXIMUM ENTROPY METHOD	161
THE COMPUTATIONAL ALGORITHM	164
SOME EXAMPLES OF IMAGES RESTORATION	168
CONCLUSION	173
ACKNOWLEDGMENTS	174
REFERENCES	175
PART VIII. LOCAL FEATURE ENHANCEMENT OF SYNTHETIC RADIO IMAGES BY ADAPTIVE KALMAN FILTERING	177
ABSTRACT	178
INTRODUCTION	179

PRELIMINARY DEFINITIONS	182
IMAGE SEGMENTATION BY LOCAL SPATIAL ACTIVITIES	184
SOURCE STRUCTURE AND NOISE DISTURBANCE MODELING	188
KALMAN FILTERING IMAGES	192
ENHANCING RADIO SYNTHETIC IMAGES NGC 7027 AND 3C47	196
CONCLUSION	203
ACKNOWLEDGMENTS	204
REFERENCES	205
PART IX. T-CALIBRATION: A NEW TECHNIQUE FOR CORRECTING ATMOSPHERIC-INDUCED PHASE ERRORS OF A SYNTHETIC-APERTURE ANTENNA ARRAY BY TIME SERIES MODELING AND KALMAN FILTERING	208
ABSTRACT	209
INTRODUCTION	211
INTERFEROMETER AND ATMOSPHERIC DISTURBANCE	214
MODELING ALGORITHM	217
Parameter Estimation with Unknown Signal and Knowledge of Noise Covariances	219
Parameter Estimation with Unknown Noise and Knowledge of Signal Covariances	221
Parameter Estimation with Unknown Signal and Knowledge of Noise Variance	222
PARAMETER ESTIMATION COMBINING THE TIME SERIES MODELING METHOD WITH OTHER DATA PROCESSING ALGORITHMS	223
Parameter Estimation with Data from Calibrator Sources	223
Parameter Estimation with CLEAN Components	224
PHASE ESTIMATION BY KALMAN FILTER	226
SOME EXPERIMENTAL RESULTS OF T-CALIBRATION	229
A Simulation of Filtering Antenna Gains	229
Correcting Baseline-Based Phase Data	232
A Simulation of Filtering Phase Data Combining with the CLEAN	235
SUMMARY	240
ACKNOWLEDGMENTS	243

REFERENCES	244
GENERAL SUMMARY	247
ACKNOWLEDGMENTS	250

GENERAL INTRODUCTION

The principal topic of this dissertation is the development and application of signal and image processing to Nondestructive Evaluation (NDE) and radio astronomy. Given NDE radiographic images and radio astronomy interferometer data (or images), some common demands are: noisy data are required to be analyzed and filtered, weak features are needed to be enhanced and detected, blurred details are wanted to be reconstructed and restored. In this work, some algorithms and techniques are proposed and developed for these needs.

The dissertation consists of nine parts which are related to nine papers published or submitted for publication. Each of them has a specific and unique topic related to signal processing or image processing in NDE or radio astronomy. These topics are: interferometer data analysis and time series modeling; industrial NDE X-ray radiographic image analysis and feature extraction; adaptive and edge preserving filtering for NDE images; image segmentation by a rule based expert system; NDE X-ray image enhancement by Kalman filtering; deblurring NDE X-ray and infrared images by maximum entropy deconvolution; local feature enhancement for radio astronomy images; and a new technique for interferometer phase data correction.

Data analysis and time series modeling of phase data from the Very Large Array (VLA) (operated by NRAO¹) are introduced in Part I (J. P.

¹The National Radio Astronomy Observatory is operated by Associated University Inc., under contract with the National Science Foundation.

Basart and Y. Zheng. 1986. Modeling very large array phase data by the Box-Jenkins method. *Radio Science*, 21: 863-881). Phase fluctuations caused by atmospheric disturbance place a limitation on the quality of radio astronomy images made by interferometers. As the baselines and observation frequency increase, phase variation erratically increases. Phase errors produce blurred and noisy images of sources observed. Many studies have been made of the behavior of the atmospheric phase fluctuation. We found that the phase fluctuation is correlated in time and that it can be described by time series ARIMA models. ARIMA models of the phase variation provide good short-term forecasting of the phase in time.

The factors degrading the quality of the NDE X-ray radiographic images are analyzed in Part II (Y. Zheng and J. P. Basart. 1988. Image analysis, feature extraction, and various applied enhancement methods for NDE X-ray images. In D. O. Thompson and D. E. Chimenti, eds. *Review of Progress in Quantitative NDE*, Vol. 7 (in press) Plenum Press, New York.). The analysis focuses on image contrast, unsharpness, and film graininess. Then, various enhancement techniques are applied to improve the quality of the images. They are: histogram equalization, background trend removal, median filter, sigma filter, adaptive smoothing filter, and the Kalman filter. Feature extraction was done by detecting flaws with image segmentation and by detecting flaw spatial activity with the modified masking function.

The application of the Kalman filter in NDE X-ray image enhancement is introduced in Part III (J. P. Basart, Y. Zheng and E. R. Doering. 1987. Application of adaptive regional Kalman filter to X-ray images in NDE. 6: 767-772 in D. O. Thompson and D. E. Chimenti, eds. Review of Progress in NDE. Plenum Press, New York.). Image segmentation was done to find wide-sense stationary regions of an image. The parameters of autoregressive models of image processes were estimated with a white noise model. A Kalman filter adapted to the local features was used to enhance the images.

A rule-based expert system was developed for a robotic scheme to segment and enhance an image. It is introduced in Part IV (Y. Zheng and J. P. Basart. 1987. Automatic image segmentation, modeling and restoration with a rule-based expert system. pp. 421-425 in D. M. Etter, ed. Twentieth annual asilomar conference on signal, system, and computers. IEEE. Inc., New York.). A tree structure search algorithm was used. Knowledge is represented as the condition-action rules. Segmentation and noise variance estimation were realized by a set of condition-action rules. The whole enhancement procedure was governed by a set of control rules.

A modeling and Kalman filtering scheme reducing signal dependent colored noise was developed for enhancing NDE X-ray radiographic images. It is given in Part V (Y. Zheng and J. P. Basart. 1988. NDE X-ray image modeling and adaptive filtering considering correlated noise. In D. O. Thompson and D. E. Chimenti, eds. Review of Progress

in Quantitative NDE. Vol. 7 (in press) Plenum Press, New York.). In this work, a technique is developed to remove noise fluctuations caused by film-grain noise, quantum fluctuation, and film dirt on the images. Some knowledge of a radiographic X-ray image forming mechanism is employed to obtain AR representations for both image and noise processes. The models are transformed to state-space forms so that the Kalman filter can be used to separate processes and enhance an image.

Computational problems of Kalman filtering VLA phase data are reported in Part VI (Y. Zheng and J. P. Basart. 1988. Kalman filtering VLA phase data with a supercomputer. pp. 141-148 in T. J. Cornwell ed. The Use of Supercomputers in Observational Astronomy. NRAO/AUI, Green Bank, WV). The computational analysis was done for the Kalman filter program. A comparison of performance was made for a DEC-10 computer and a CRAY X-MP/24 computer. Some optimization and vectorization methods were introduced. Statements supporting the purchase of a supercomputer by NRAO were made.

Part VII introduces a practical maximum entropy method for deblurring NDE X-ray and infrared images (Y. Zheng and J. P. Basart. 1987. Deblurring NDE X-ray images and infrared images by maximum entropy method. presented at the Twenty-First Annual Asilomar Conference on Signal, System, and Computers, Pacific Grove, CA.). A fast and practical MEM was studied and developed for NDE images. This MEM uses a Newton-Raphson approach to find a conditional extremum vector of the entropy equation subject to conditions enforced by

Lagrange multipliers. The consistency between the observed data and the reconstructed data is measured by the χ^2 statistic. Only two FFTs are required in each iteration which are less than many other MEM methods.

Part VIII represents an adaptive Kalman filter scheme for enhancing radio astronomy images (Y. Zheng and J. P. Basart. Local feature enhancement of synthetic radio images by adaptive Kalman filtering. In preparation for submittal to Astronomical Journal for publication.). This technique focuses on local feature enhancement which is different than other radio astronomy image restoration methods. The enhancement is specially effective in the low signal to noise region where the intensity of source and noise are at the same order and one can not confidently defined a source structure.

Part IX introduces a new technique for interferometer phase corrections (Zheng, Y., and J. P. Basart, and Y. S. Koh. T-calibration: a new technique for correcting atmospheric-induced phase errors of a synthetic-aperture antenna array by time series modeling and Kalman filtering. In preparation for submittal to Radio Science for publication). This new technique for estimating phase data by time series modeling and Kalman filtering is combined with other radio astronomy data processing algorithms such as CLEAN and calibration.

PART I. MODELING VERY LARGE ARRAY PHASE DATA BY THE BOX-JENKINS METHOD

ABSTRACT

The quality of radio astronomical images made with an antenna array depends upon atmospheric behavior. As baselines and frequencies increase, phase variations become increasingly erratic. The phase fluctuations are time dependent and we found them to be correlated in time order in each baseline. We can represent these correlations by stochastic models. Models obtained by the Box-Jenkins method are referred to as autoregressive integrated moving average processes (ARIMA). ARIMA models of VLA (operated by NRAO²) phase provide good short-term predictions that may be useful for improving present calibration techniques. ARIMA models of VLA phase are data dependent and can be used in a variety of situations. A technique that works in all cases can be programmed into a software package such that modeling can be accomplished with no operator interactions. Another important application of ARIMA models involves the use of Kalman filtering to reduce the atmospheric effects when self-calibration does not work well. The performance of the Kalman filter critically depends upon the models of the processes. An ARIMA model of the phase fluctuation can be represented in a state space form as noise in the Kalman filter equations.

²The National Radio Astronomy Observatory is operated by Associated Universities Inc., under contract with the National Science Foundation.

INTRODUCTION

Atmospheric disturbances place a severe limit on the quality of maps made with high-resolution radio astronomy interferometers. Instruments such as the very large array (VLA) [Thompson et al., 1980; Napier et al., 1983] have reduced performance, especially at long baselines and high frequencies, when there is differential atmospheric behavior between antenna line-of-sight paths to the celestial radio source. The study of atmospheric disturbances and how to correct for them is a continuing process.

It has long been known that atmospheric water vapor is a strong contributor to phase fluctuations in radio interferometers operating at centimeter wavelengths [Baars, 1967; Hinder, 1970; Wesseling et al., 1974]. As radio interferometers increased in baseline length and in frequency of operation, the number of atmospheric studies increased. These are well represented by the following authors: Mathur et al. [1970], Hinder and Ryle [1971], Hargrave and Shaw [1978], Hamaker [1978], Dravskikh and Finkelstein [1979], Han [1980], Moran and Rosen [1981], and Armstrong and Sramek [1982].

Correcting the data for atmospheric phase changes is normally accomplished by two calibration procedures. The first procedure (empirical calibration) is to collect data periodically during the observation of a strong pointlike calibrator source located as near as possible (in an angular sense) to the unknown source. From the calibrator data, correction factors are determined and applied to the

unknown source data. Correction factors are often determined by either an interpolation between calibration points or by averaging several calibrator points together. Under reasonable atmospheric conditions this method accounts for most of the long-term amplitude and phase variations. It cannot be perfect because the calibrator source is observed at a different position and at a different time than the program source.

The second calibration procedure involves a scheme to further calibrate the data by using only the program source data (no calibrator source data). This method is often called "self-calibration" [Readhead and Wilkinson, 1978; Schwab, 1980]. The self-calibration method grew out of hybrid mapping techniques developed in VLBI by Rogers et al. [1974] and Readhead and Wilkinson [1978]. The fundamental concept behind the hybrid technique is that phase closure is free of instrumental and atmospheric effects. The concept of phase closure was discovered by Jennison [1958]. In Readhead and Wilkinson's scheme, for N antennas, $N-1$ estimates of the true phase are derived from an initial model for the first iteration or from CLEAN components on subsequent iterations. Then, all remaining phases are obtained by solving $N(N-1)/2$ closure phase relations. The CLEAN components are produced by the CLEAN deconvolution algorithm developed by Hogbom [Hogbom, 1974; 1984; Schwarz, 1978; Clark, 1980; Cornwell, 1983]. The CLEAN algorithm is a very efficient method for eliminating the sidelobes in a synthesized radio image. The self-calibration developed by Schwab

finds correction factors (antenna gains) by minimizing the mean-square error between pseudo data generated with the CLEAN components and the actual data. This method is better than the hybrid technique since the solutions are obtained by an iterative nonlinear least-squares method based on the least squares principle. Schwab's method often produces very good improvement in the data calibration. This gives a lower residual noise level in maps and increases the dynamic range of maps limited by phase noise rather than by the noise temperatures of the receivers. The method works well when there is a point-like object in the field-of-view or if there is a good model of an extended source.

With the successes mentioned above it is natural to turn next to improving calibration procedures for more severe atmospheric conditions such as those occurring for summer time water vapor, long baselines, and high microwave frequencies.

One approach to this problem is to measure the atmospheric path length fluctuations along each antenna's line of sight with a microwave water vapor radiometer (WVR), and then find the differential path length change between every antenna pair. The difference can be used as a correction factor. This technique has had some success [Resch et al., 1984]. However, as pointed out by Resch et al., there is a problem to be resolved before the procedure can be used routinely. The atmospheric brightness is not perfectly related to path length with the present opacity algorithm. This may cause the prediction of the WVRs to fail under conditions of heavy clouds containing drops whose average size is comparable to a wavelength [Westwater and Guiraud, 1980].

Since deterministic methods have difficulties, we are testing an alternate method. Our approach to this problem is to apply the theory of Kalman filtering [Kalman, 1960; Kalman and Bucy, 1961]. Since its inception the Kalman filter has been extensively used in both control systems and in signal-processing systems [Brown, 1983]. Okatan and Basart [1979] applied the Kalman filter to 26 MHz VLBI data on 950 km baseline and significantly reduced the phase fluctuations caused by the ionosphere.

The primary feature of the Kalman filter that is useful to us is the mathematical formalism into which we can put all the information about the atmosphere, the radio telescope, and the stochastic nature of the signal and noise. The filter can distinguish between different stochastic processes (even if their spectra strongly overlap). It estimates the variables of interest in a recursive and optimal way. The primary limitations are the complexity of the mathematical formulation and the computation time. Another feature which may be useful in the calibration of very long baseline interferometry (VLBI) data is the recursive nature of the Kalman algorithm.

Kalman filter theory is couched in state-variable theory which is the modern way to analyze control systems. To minimize computation time it is important to minimize the number of descriptors of the data and of the system. Hence, our approach is to use the simplest models that can do a productive job in calibration.

The success of the Kalman filter in reducing unwanted effects in data depends critically upon the models describing the data and the system. In a sense, the Kalman filter is "tuned" to a particular process of interest. If the filter is properly tuned by correct modeling, the results will be equal to or (hopefully) superior to conventional processing techniques. However, if the filter is improperly tuned by choosing incorrect models, the results can easily be worse than the original data.

Because of the acute sensitivity of the Kalman procedure to the model, we must devote considerable time to studying the characteristics of the data in order to properly describe them. In this paper we will discuss the characteristics of noisy phase data collected with the VLA and how we represent the data by stochastic models. Since the effect of phase errors dominate the effect of amplitude errors in the type of data we are studying, we are concentrating our efforts on phase at this time.

Data used in this study were collected by R. A. Sramek at the VLA in an ongoing program in which he regularly analyzes atmospheric disturbances (see, for example, Armstrong and Sramek [1982]). Typically, analysis procedures presented in the literature by various authors concentrated on the statistics of data gathered and combined over several observations. Since we are attempting to correct the data as a function of time, we have taken a different approach. Our studies have focused on the stochastic behavior of the time series of phases

over one observing run or fraction thereof. In this paper we report the statistical behavior of these time series and our procedure for modeling the series by the Box-Jenkins method [Box and Jenkins, 1976]. In a future paper we will report our procedure for placing the models into a Kalman filter algorithm and our method for reducing noise in the maps.

DATA CHARACTERISTICS

Our overall approach in developing a new calibration procedure for badly corrupted data, has been to start with well behaved data which can be easily calibrated by current techniques and to then proceed to more complex data. We initially studied the characteristics of data whose phase change is at most a few tens of degrees over 1 or 2 hours. Later we studied phase data with larger fluctuations which are the type of data our procedure is intended to correct.

The VLA data bases studied are listed in Table 1.1. Over 300 time series, each on a different geometrical baseline, were available in each data base. We studied a total of 184 time series in the eight data bases. Two IF channels were available for each antenna pair, but we only studied one of the pairs because the dominant atmospheric disturbances were the same on each IF channel. An antenna located near the center of the array was used as a phase reference in each time series. The rms values of the phase in the time series varied from a few degrees to over 40° .

Figures 1.1, 1.2 and 1.3 illustrate time series for phase data with various amounts of fluctuation. The term "phase" as used here is the phase difference between the signals from two antennas. If all geometric aspects are properly accounted for, as assumed here, the ideal phase will be a constant since we have observed a point source. The deviations from a straight line are a measure of perturbations, such as caused by the atmosphere and the electronic system.

Table 1.1. VLA Data Analyzed

Date	Radio Source	Configuration	Wavelength, cm	Record Length, min	Average Time, sec	Number of Series Analyzed
Dec. 13, 1980	0316+161	A	6	204	60	22
Dec. 17, 1980	1741-038	A	6	201	60	23
Dec. 18, 1980	0316+161	A	6	206	60	21
May 7, 1982	0316+161	A	6	204	60	24
May 21, 1982	0316+161	A	6	207	60	23
May 21, 1982	1741-038	A	6	207	60	24
Nov. 13, 1982	0923+392	A	6	163	60	22
Apr. 7, 1984	3C 84	C	2	115	50	25

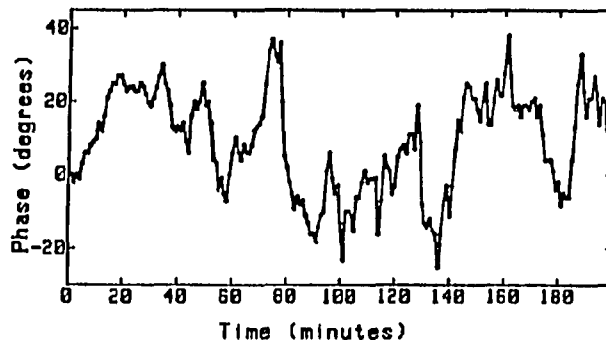


Figure 1.1. Time series of phases averaged to one-minute time intervals. Data were collected on May 21, 1982, on a 10.7 km baseline at a wavelength of 6 cm. The mean is 9° , rms is 14° , and peak-to-peak change is 63° . The radio source was 0316+161.

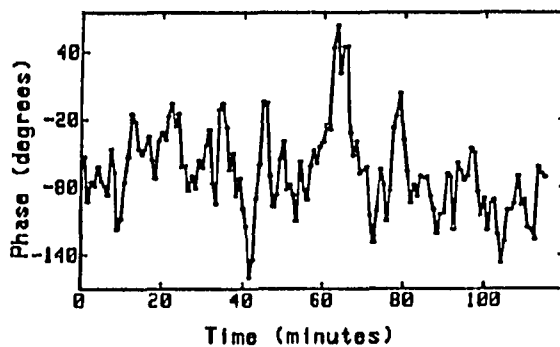


Figure 1.2. Phases measured on April 7, 1984. Integration time = 50 s, baseline = 1.09 km, wavelength = 2 cm, mean = -66° , rms = 39° , peak-to-peak = 224, and the source was 3C 84.

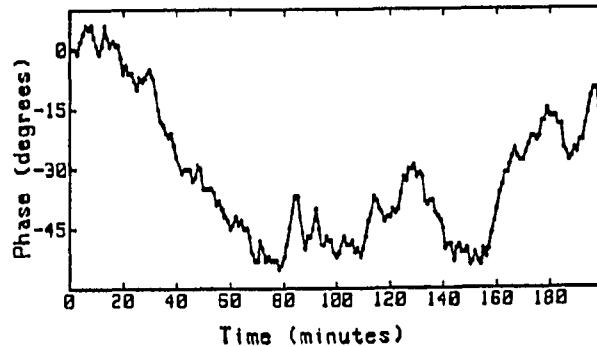


Figure 1.3. An example of a time series nonstationary in the mean. The data were collected on May 21, 1982, on a 21.22 km baseline at a wavelength of 6 cm. The integration time = 60 s. The mean = -32° , rms = 17° , and peak-to-peak = 61° . The radio source was 1741-038

Essentially all of the significant phase perturbations displayed in this paper are assumed due to the atmosphere. Figure 1.1 shows a relatively small variation which would be satisfactorily handled by the usual calibration procedures. The large fluctuations shown in Figure 1.2 are typical of the fluctuations we are trying to correct. Figure 1.3 illustrates a nonstationary time series. This type of large nonstationarity has appeared mainly in data with relatively slow fluctuations. We have defined a numerical quantity to help describe the type of variation in a time series. Variance describes the magnitude of the variations, but it does not describe how fast these

variations occur. To describe the combination of large swings with the rate of variation we use the first difference,

$$d_{ij}(t) = \phi_{ij}(t) - \phi_{ij}(t-1) \quad (1.1)$$

where $\phi_{ij}(t)$ is the phase at time t for baseline ij and $t-1$ is one sampling time interval before t . Taking the expectation of the absolute value of d and calling it A_d gives

$$A_d = E[|d_{ij}(t)|] \quad (1.2)$$

The faster the phase variation is, the larger is A_d . For Figures 1.1, 2.1, and 3.1, A_d is 4.2° , 22.1° , and 1.9° , respectively. For the data sets we have studied, A_d varies from 4° to 24° for 2-cm wavelengths and from 2° to 8° for 6-cm wavelengths. Whereas the rms varies from 5° to 42° for 2 cm and from 3° to 36° for 6 cm. There is not much difference between the maximum values of rms at the two wavelengths, but A_d is considerably more at 2 cm than at 6 cm. The ratio of the maximum A_d (24:8) is equal to the inverse of the wavelength, as we would expect. In the data sets we analyzed, 2-cm phase data varies much more rapidly than 6-cm data and the rms measurements do not obviously indicate this difference.

As shown in Figure 1.3, the large rms values of 6-cm data are caused by the large drifts. These drifts are very slow, as the A_d show. A_d are related to the average phase change in each time interval. A_d are nearly constant and the largest difference in them

does not exceed 3° in each 6-cm data set. On the other hand, the rms values of 6-cm data do not monotonically increase as the baselines increase. One possible interpretation is that these phase variations of 6-cm data may have significant contributions from system noise and instabilities in addition to the dominant atmospheric disturbances. Another possibility is that the cell sizes of the water vapor are much less than the longest baselines for our data sets. The mismatch between rms and real phase fluctuation is due to the nonstationarity of the phase time series. In this case, A_d gives a better measurement of atmospheric phase fluctuation. The square of the A_d values of 2-cm data vary approximately as baseline length to the 0.88 to 1.14 power. This range of power overlaps the low end of the power of the phase structure function found by Armstrong and Sramek [1982], and is a little less than the $5/3$ power of the phase structure function for the Kolmogorov spectrum [Ishimaru, 1978, p. 380]. The structure function can be written in the form

$$D(b) = c_\phi^2 b^{a-2} \quad (1.3)$$

where b is the baseline length and c_ϕ^2 is a constant. Armstrong and Sramek find a to lie in the range 3.1 to 3.8 for baseline lengths ranging from 1 to 10 km. This is consistent with the a for the short term temporal power spectrum of precipital water vapor which ranges from 3.0 to 3.2 [Hogg et al., 1981].

Generally, phase variations increase with baseline since the atmosphere becomes more dissimilar between antennas. Phase plots for three baselines at 2-cm wavelength are shown in Figure 1.4. As we would expect, the longest baseline has the largest variations. A_d for these plots is 18.5° , 6.4° , and 2.8° for series A, B, and C, respectively.

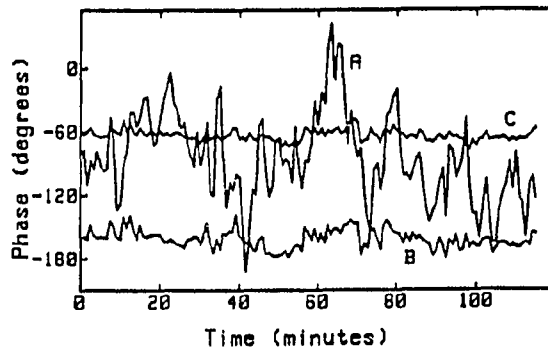


Figure 1.4. Time series for three baselines. Observation was on April 7, 1984, with wavelength = 2 cm. The source was 3C 84. The baselines were 1.98 km for series A, 0.18 km for series B, and 0.09 km for series C. The mean values for the series are -83° , -159° , and -63° , respectively. The rms values are 43° , 16° , and 10° , respectively

Combining baseline and time information we can present the data in histogram form. Two sets are shown here in Figures 1.5 and 1.6. Since phase is relative and some reference must be chosen, the phases from

the various baselines were forced to agree at the beginning of the time series. The first 50 min of data for a 200-min run at 6-cm wavelength are shown in Figure 1.5. The mean is near zero for all baselines and the dispersion increases with baseline. The rms values vary from 2.6° at 0.48 km to 11.5° at 21.01 km. In Figure 1.6 we show the last 50 min of the 200-min run. The means are now deviating about zero and the dispersion has increased. The rms values vary from 2.9° at 0.48 km to 15.7° at 21.01 km. The variation of the mean and rms with time indicate nonstationary properties of the phase. Although the means appear to be scattered in time, the relatively small phase fluctuations can be corrected by standard techniques. The phase variations of the data sets we have studied do not show a consistent behavior as a function of baseline on a short time basis for the 6-cm data sets. Roughly, the variation increases with baseline most of the time, but very often some intermediate baselines have less phase fluctuation than some of the shorter baselines.

For large phase fluctuations at 2 cm, the behavior is different. The rms is large and nearly stationary. Histograms for a range of baselines are shown in Figure 1.7. This is for the first 25 min of the run. The shape of the histograms, as a function of baseline, for the last 25 min of the 115-min run are very similar to those for the first 25 min (not shown). The mean changed systematically for all baselines by about 50° over the 115 min.

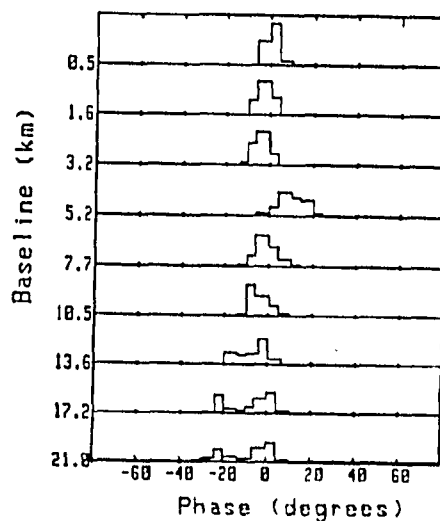


Figure 1.5. Time-space histograms of phase data. Each histogram contains the first 50 min of a 200-min run on the southeast arm of the VLA on May 21, 1982. Phases were set to zero at the beginning of the run. The wavelength was 6 cm and the source was 0316+161

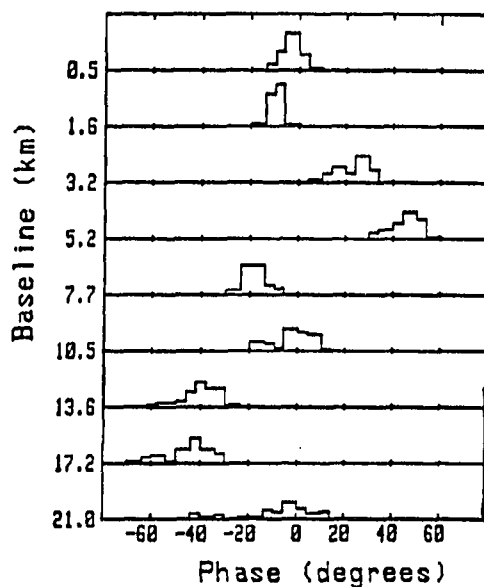


Figure 1.6. Time-space histograms for the last 50 min of the run described in Figure 1.5

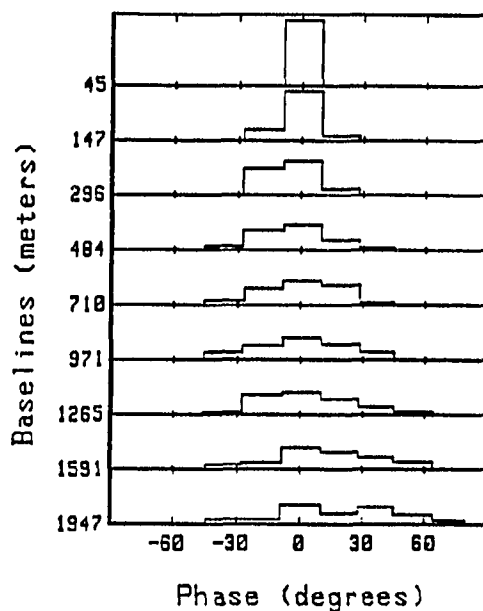


Figure 1.7. Time-space histograms for very noisy phase data. Each histogram contains 30 min of data collected on the southwest arm of the VLA on April 7, 1984. The wavelength was 2 cm and the source was 3C 84

An example of a well-defined trend in phase fluctuation vs baseline is shown in Figure 1.8. The three curves with tick marks correspond to the three arms of the VLA. The curves follow a power law up to a baseline length of 2 km. A least squares fit to the three curves gives

$$\text{rms}(b) = (1.510 \pm 0.005)b^{0.44 \pm 0.005} \quad (1.4)$$

where b is the baseline length in kilometers. The rms of the residuals is 5.70° . The exponent of b is similar to the exponent in the power law of A_d . The plot of (1.4) is given by the curve with no tick marks in Figure 1.8 and closely follows the data.

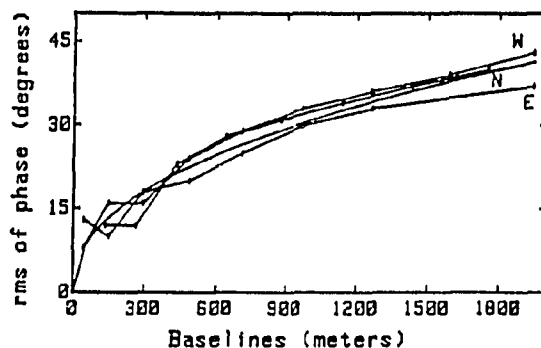


Figure 1.8. RMS plots of phase vs. baseline. N, W, and E refer to the north, southwest, and southeast arms of the VLA, respectively. The data were collected on April 7, 1984, while observing the radio source 3C 84 at 2cm

Another routine function we used to analyze the behavior of phase in time order was the autocorrelation function. It plays a significant role in our method of modeling the time series. Figures 1.9 through 1.12 illustrate the various types of autocorrelation functions for the various types of observed phase behavior. One of the most typical behaviors is shown in Figure 1.9 in which the autocorrelation slowly decreases to nearly zero. It is a Gauss-Markov process which can be modeled by the function

$$R(\tau) = \sigma^2 e^{-\beta\tau} \quad (1.5)$$

The parameter values fitted to (1.5) are $\sigma^2 = 1.00$ and $\beta = 0.141 \pm 0.0005$. β was eliminated from the first 25 autocorrelation coefficients by a least-squares method. The rms of the residuals is 0.0352. The solid line without dots in Figure 1.9 was generated from (1.5). Occasionally the autocorrelation appears as a pseudoperiodic function as shown in Figure 1.10. It can be modeled by

$$R(\tau) = \sigma^2 e^{-\beta\tau} \cos \omega_0 \tau \quad (1.6)$$

which is a periodic Gauss-Markov random process. The parameter values are $\sigma^2 = 1.00$, $\beta = 0.086 \pm 0.005$ and $\omega_0 = 27.88 \pm 0.005$. β and ω_0 were estimated from the first 22 autocorrelation coefficients by a least squares method. The rms of residuals is 0.0333. The solid line without dots in Figure 1.10 is a plot of (1.6).

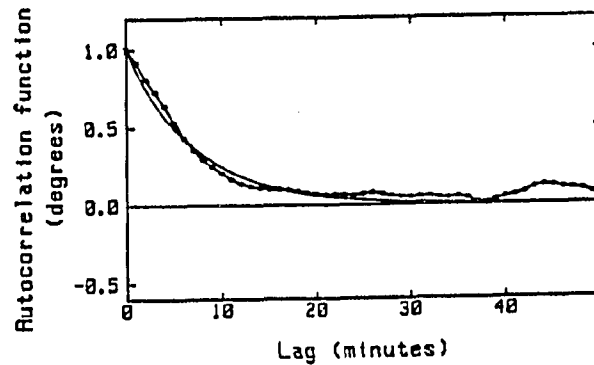


Figure 1.9. An autocorrelation function typical of many time series of the VLA phase. The data are the same as that in Figure 1.1

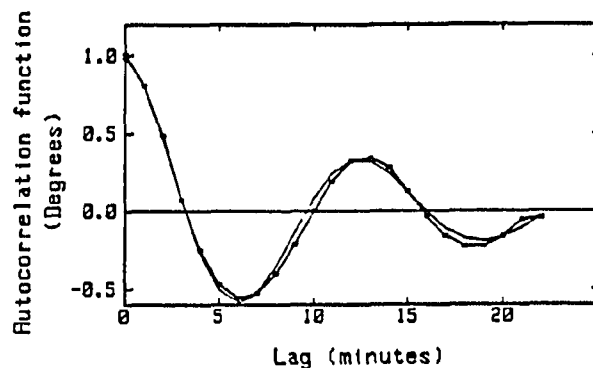


Figure 1.10. A pseudoperiodic autocorrelation function of the VLA data. The data were collected while observing source 0316+161 on March 7, 1982, at 6-cm wavelength

An example of phase variation that resembles white noise is shown in Figure 1.11. The rapid drop near zero lag and the subsequent near zero values indicate that the noise sequence is nearly uncorrelated. Its model is given by

$$R(\tau) = \delta(t) + 0.311\delta(\tau-1) \quad (1.7)$$

The final example of autocorrelation functions is given in Figure 1.12. The very long decay is indicative of a nonstationary mean. The modeling process for this is discussed later.

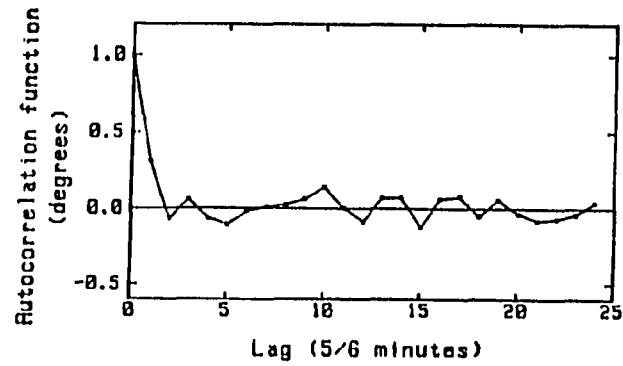


Figure 1.11. An autocorrelation similar to an ideal delta function. The data are from the same database but a different baseline (0.76 km) described in Figure 1.2

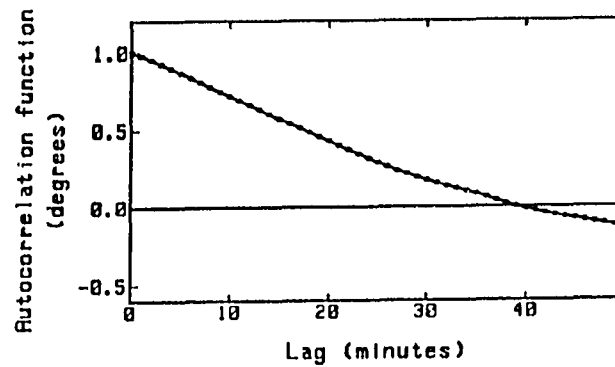


Figure 1.12. An autocorrelation function of a nonstationary time series

TIME-SERIES MODELS OF THE PHASE

As shown in the previous section, atmospheric phase variation is correlated in time order. In this section we will show that the phase variation can be represented by a stochastic model. Consequently, our modeling of the VLA phase is oriented toward a time-series approach. This also is the order in which the data are collected. For mapping, the data must be reordered by spatial coordinates which scrambles the atmospheric information. Complete modeling would include spatial variation in addition to temporal variation. However, spatial variation has been omitted for simplicity. It is clear from sample studies that adjacent antennas have similar phase behavior, but this coupling is omitted also. Spatial variations and correlations between antennas will be considered in the future. The objectives of this modeling procedure are to (1) find difference equations to describe phase propagation error due to the atmosphere, (2) statistically test whether the models represent the stochastic aspects of the VLA phase data, (3) forecast atmospheric variation, and (4) determine the VLA phase models to be placed in Kalman filters to reduce phase errors.

There are two ways to model a system: an internal description and an external description. The internal description is based on the knowledge of the internal structure of a system. The external description is based on the behavior of the input and output of a system. Limited by contemporary technology at the present time, a practical useful physical model of atmospheric variations based on the

knowledge of the internal structure of the atmosphere, cannot be built yet. The atmosphere is a "black box" in a sense. Clearly, what we can do is extract descriptive properties of the atmospheric phase disturbance from observations. Then we build models that adequately describe the time variation of the phase.

In the previous section, samples of data and models for the data were given. The four samples required four different models whose mathematical forms were obtained by inspection. If one is to automate the modeling in a computer, it is desirable to minimize human interaction. That is, one should not have to manually study the autocorrelation of each series and then tell the computer what model to use. A procedure is required which is capable of satisfactorily modeling as many different types of phase behavior as feasible. The procedure we have chosen which has proven to be satisfactory thus far is Box-Jenkins time series analysis. We will give a brief description of it here before discussing our models of the VLA phase data. For more details, the reader is referred to the texts by Pankratz [1983], Anderson [1976], McCleary and Hay [1980], and the master text by Box and Jenkins [1976]. Three other useful references with many applications are Anderson [1982a, b, 1983].

The idea behind the Box-Jenkins approach is that a measured point in a time series may be statistically related to other measured points in the series, and the goal in the Box-Jenkins analysis is to find a good way of stating this statistical relationship. Thus, the Box-

Jenkins approach is based on modeling the correlation structures. A variety of ARIMA models obtained by the Box-Jenkins approach implies a variety of different correlation structures [Pankratz, 1983]. The major tools in the Box-Jenkins approach are autocorrelation functions (ACF) and partial autocorrelation functions (PACF). The model estimation is often based on a nonlinear least squares (NLS) procedure. Compared with other traditional time-series analyses, the Box-Jenkins approach has three advantages:

1. The concepts associated with ARIMA models are derived from a solid foundation of classical probability theory and mathematical statistics. Many other methods are derived in an intuitive way.
2. ARIMA models are a family of models. According to a certain strategy, one or more appropriate models can be selected out of this family.
3. An appropriate ARIMA model produces optimal predictions which have a small mean-squared forecast error.

The Box-Jenkins method consists of three main steps: identification, estimation, and diagnostic checking. Identification is the process of choosing the functional form of a model. Estimation is the process of determining the parameters of the model, and diagnostic checking is the testing of the model against the data to see if the residuals are "white" and uncorrelated. If the test results are not satisfactory, the entire procedure is repeated.

Model identification for a stationary time series starts with finding the autocorrelation and partial autocorrelation coefficients. Normalized autocorrelation coefficients are given by the standard formula

$$r_k = \frac{\sum_{t=1}^{N-k} (z_t - \bar{z})(z_{t+k} - \bar{z})}{\sum_{t=1}^N (z_t - \bar{z})^2} \quad (1.8)$$

where N is the number of points in the sample series, z_t is a sample at time t , and \bar{z} is the average value of the series. Partial autocorrelation coefficients are found recursively using

$$p_{11} = r_1 \quad (1.9)$$

and

$$p_{kk} = \frac{r_k - \sum_{j=1}^{k-1} p_{k-1,j} r_{k-j}}{1 - \sum_{j=1}^{k-1} p_{k-1,j} r_j} \quad (1.10)$$

where

$$P_{kj} = P_{k-1,j} - P_{kk}P_{k-1,k-j}$$

$$k = 2, 3, \dots$$

$$j = 1, 2, \dots, k-1$$

The estimated ACF and PACF are measurements of statistical relationships within a data series. ACF measures the joint probability distribution of pairs of values (z_t, z_{t+k}) of the time series, separated by a constant interval or lag k . PACF measures the relation of (z_t, z_{t+k}) taking into account the effect of pairs of (z_{t+1}, z_{t+k}) , $(z_{t+2}, z_{t+k}) \dots (z_{t+k-1}, z_{t+k})$. PACF exploits the fact that where an $AR(p)$ (auto-regressive process of order p) process has an ACF infinite in extent, it can be described in terms of p nonzero coefficients of the ACF. For an $AR(p)$, the coefficients of PACF will be non-zero for k less than or equal to p and zero for k greater than p . The estimated ACF and PACF are guides in choosing one or more ARIMA models that seem appropriate. The basic idea is: every ARIMA model has a unique theoretical ACF and PACF associated with it. At the identification stage we compare the estimated ACF and PACF calculated from the data with various theoretical ACFs and PACFs. The model, whose theoretical ACF and PACF most closely resemble the estimated ACF and PACF of the data, will be chosen as a candidate model. The general form of ARIMA models can be represented by the autoregressive process

$$z_t = C + \phi_1 z_{t-1} + \phi_2 z_{t-2} + \dots + a_t \quad (1.11)$$

and the moving average process

$$z_t = C - \theta_1 a_{t-1} - \theta_2 a_{t-2} - \dots + a_t \quad (1.12)$$

where z_t is the current data point, C is a constant, and the ϕ 's and θ 's are unknown coefficients to be found at the estimation stage. The noise term a_t is often called a random shock. It also is a residual of the fit of the equation to the time series. These representations for the current data point are written in terms of the previous data points in the former case, and in terms of the previous random shocks in the latter case.

A more general representation for these series is given by

$$\phi(B) (1-B)^d \tilde{z}_t = \theta(B) a_t \quad (1.13)$$

where B is the backward shift operator defined by $Bz_t = z_{t-1}$ and $\tilde{z}_t = z_t - \bar{z}$. $\phi(B)$ and $\theta(B)$ are functions of B and defined by

$$\phi(B) = 1 - \phi_1 B - \phi_2 B^2 - \dots - \phi_p B^p$$

and

$$\theta(B) = 1 - \theta_1 B - \theta_2 B^2 - \dots - \theta_q B^q$$

The d power of $(1-B)$ in (1.13) arises because many time series are not stationary in the mean. Taking the first or second difference of the series usually makes the series stationary. The value of d is the order of the difference required to make the series stationary in the mean. The notation $ARIMA(p,d,q)$ represents the order of the process.

Before a time series can be modeled by an ARIMA process, it must meet several requirements. The data should be (1) sampled at equally spaced time intervals, (2) autocorrelated and (3) stationary in variance and in the mean. Additionally, for N data points, the maximum order of a model lies in the range $N/4 \sim N/3$. If requirement one is not met, the data must be resampled at equal intervals. If the data are uncorrelated, the process is white noise which would be characterized by its variance. If the series is not stationary in the mean, one or more differences must be taken. Usually one is enough in our situation. If the series is not stationary in variance, a nonlinear transformation must be made on the series. The selection of a transformation function is difficult to automate. However, in many cases the problem can be avoided by dividing the original series into one or more series, each of which is nearly stationary in variance.

The modeling process starts with a uniformly spaced stationary time series. The ACF and PACF are calculated. From these one can determine the form of the model to choose. An example of data that easily fit the Box-Jenkins modeling procedure is shown in Figures 1.13 and 1.14. The time series for this data is shown in Figure 1.2. This is an AR(1) process (autoregressive series with z_t depending upon z_{t-1} only). This information is conveyed by the exponential decay of the ACF and the one spike in the PACF. The threshold of significance is represented by the dashed lines in the figures representing a t -test value of 1.96. An AR(2) process would have an exponentially decaying

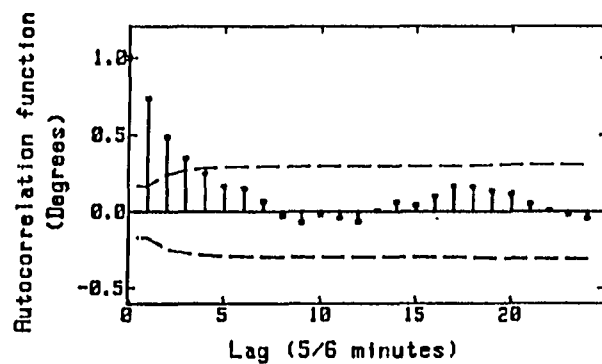


Figure 1.13. Autocorrelation of the time series shown in Figure 1.2. The dashed line represents a level of significance given by a t-test value of 1.96

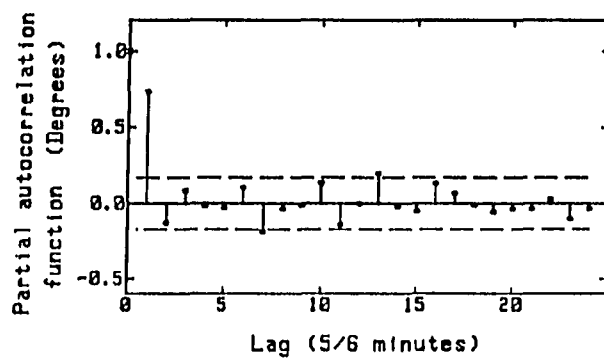


Figure 1.14. Partial autocorrelation of the time series shown in Figure 2. The dashed line represents a level of significance given for a t-test value of 1.96

ACF and two significant spikes in the PACF and so on for higher order AR processes. With a suggested model in this case of

$$z_t = C + \phi_1 z_{t-1} + a_t \quad (1.14)$$

we next must find ϕ_1 and C . The estimation of the parameter, ϕ_1 , is carried out on the computer with a maximum likelihood (ML) or least squares (LS) approach [Box and Jenkins, 1976]. Since the likelihood function from which ML estimates are derived reflects all useful information about the parameters contained in the data, the results of the ML approach have attractive statistical properties. However, the computation of the ML approach is cumbersome and requires a lot of computer time. On the other hand, if the a_t are normally distributed, then LS estimations are either exactly or very nearly ML estimates [Box and Jenkins, 1976]. Then, LS is recommended in practice. Since the set of equations generated by minimizing the sum of squared residuals is nonlinear, the LS estimates of an ARIMA model require the use of a nonlinear least squares (NLS) method. The method which we used is called "Marquardt's compromise" [Marquardt, 1963; Box and Jenkins, 1976]. Marquardt's method is a combination of two NLS procedures: Gauss-Newton linearization and the gradient method [Pankratz, 1983]. It combines the advantages of these two approaches which not only converge to NLS estimates but also converges relatively quickly. The basic idea of Marquardt's approach is that starting with some initial values for the parameters of an ARIMA model, it finds new parameters which produce a smaller sum of squared residuals in a systematic way.

For a general ARIMA $(p,0,q)$ process describing a stationary time series with N samples, we can estimate the ARIMA coefficients with the following procedure (slightly modified and outlined from Box and Jenkins [1976]. The threshold values are ours.) 1. Find initial values of the AR parameters, ϕ_k , called ϕ_{k0} , by solving the Yule-Walker equation

$$\begin{bmatrix} \phi_{10} \\ \phi_{20} \\ \vdots \\ \phi_{p0} \end{bmatrix} = \begin{bmatrix} c_q & c_{|q-1|} & \cdots & c_{|q-p+1|} \\ c_{|q+1|} & c_q & \cdots & c_{|q-p+2|} \\ \vdots & \vdots & \ddots & \vdots \\ c_{|q+p-1|} & c_{|q+p-2|} & \cdots & c_q \end{bmatrix}^{-1} \begin{bmatrix} c_{q+1} \\ c_{q+2} \\ \vdots \\ c_{q+p} \end{bmatrix} \quad (1.15)$$

where $\{c_k: k=0, 1, \dots, p+q\}$ is an autocovariance sequence of the phase series and $\{\phi_{k0}: k=1, 2, \dots, p\}$ is a set of solutions of the initial parameters of the AR process.

2. Find initial values of MA parameters $\{\theta_{k0}: k=1, 2, \dots, q\}$ with the Newton-Raphson algorithm which calculates a dummy vector \mathbf{X}_{i+1} at the $(i+1)$ st iteration from the values of \mathbf{X}_i at the i th iteration:

$$\mathbf{X}_{i+1} = \mathbf{X}_i - \mathbf{T}_i^{-1} \mathbf{F}_i \quad (1.16)$$

where

$$\mathbf{X} = (x_0, x_1, \dots, x_q)$$

Generally speaking, the parameters estimated from step 1 and step 2 are good approximations of the results of an iterative NLS method if the model contains all ϕ_i terms between ϕ_1 and ϕ_p and θ_j terms between θ_1 and θ_q . If a model does not contain all ϕ_i terms between ϕ_1 and ϕ_p and all θ terms between θ_1 and θ_p , Marquardt's method is used to estimate the ϕ_i and θ_j . Marquardt's procedure is the following:

3. Find initial residuals with initial parameters found in step a and step 2 using

$$a_t = (z_t - \bar{z}) - \left(\sum_{i=1}^p \phi_i (z_{t-i} - \bar{z}) - \sum_{j=1}^q \theta_j a_{t-j} \right) \quad (1.18)$$

where $t = N_0+1, N_0+2, \dots, N$ and N_0 and $\max(p, q)$ and initial sum of square residuals

$$SSR_0 = \sum_{t=N_0}^N a_t^2 \quad (1.19)$$

and the initial ϕ_i is found with (1.15), and the initial θ_j is found with (1.17).

4. Find the numerical values of

$$y_{i,t} = - \frac{\partial a_t}{\partial \beta_i} \quad (1.20)$$

by

$$y_{i,t} = [a_t(\beta_1, \dots, \beta_{i+\Delta_i}, \dots, \beta_k) - a_t(\beta_1, \dots, \beta_{i-\Delta_i}, \dots, \beta_k)] / (2\Delta_i)$$

where $\Delta_i = 0.02 \beta_i$ and $\beta = (\beta_1, \beta_2, \dots, \beta_k) = (\phi_1, \phi_2, \dots, \phi_p, \theta_1, \theta_2, \theta_q)$ starting with $\beta_0 = (\beta_{10}, \beta_{20}, \dots, \beta_{k0} = (\phi_{10}, \phi_{20}, \dots, \phi_{p0}, \theta_{10}, \theta_{20}, \dots, \theta_{q0})$.

5. Form linearized equations

$$\mathbf{A} \mathbf{H} = \mathbf{G} \quad (1.21)$$

where \mathbf{A} is a $k \times k$ matrix $\mathbf{A} = \{A_{ij}\}$

$$A_{ij} = \left(\sum_{t=N_0}^N y_{i,t} y_{j,t} \right) / \left[\left(\sum_{t=N_0}^N y_{i,t}^2 \right) \left(\sum_{t=N_0}^N y_{j,t}^2 \right) \right]^{1/2} \quad \text{for } i \neq j;$$

$$A_{ii} = 1 + \pi$$

and initially $\pi = 0.0001$;

\mathbf{G} is a $k \times 1$ vector

$$\mathbf{G}' = (g_1, g_2, \dots, g_k)$$

$$g_i = \left(\sum_{t=N_0}^N y_{i,t} a_t \right) / \left(\sum_{t=N_0}^N y_{i,t}^2 \right)^{1/2};$$

\mathbf{H} is a $k \times 1$ vector

$$\mathbf{H}' = (h_1, h_2, \dots, h_k).$$

Solve (1.21) for H and then obtain the correction factor $\Delta\beta_i$ for β_i from

$$\Delta\beta_i = h_i / (\sum_{t=N_0}^N y_{i,t}^2)^{1/2}$$

Then the new parameters are

$$\beta = \beta_0 + \Delta\beta$$

where $\Delta\beta = (\Delta\beta_1, \Delta\beta_2, \dots, \Delta\beta_k)$

6. Find new sum of the squared residuals SSR_0 by (1.18) and (1.19).

7. For the case of $SSR_1 > SSR_0$, increase the parameter by a factor 10 and then return to step 5.

8. For the case of $SSR_1 < SSR_0$, if some $|\Delta\beta_i| > \epsilon_1$ where $\epsilon_1 = 0.0001$, decrease π by a factor 10 and return to step 3 with $SSR_0 = SSR_1$ and $\beta_0 = \beta$. If all $|\Delta\beta_i| < \epsilon_1$, then the estimation is done. The final results of the elements in β are the desired ϕ and θ coefficients. The maximum number of iterations of the loop through steps 5 to h is set to 25.

9. Finally, find the constant term by

$$c = (1 - \sum_{i=1}^p \phi_i) \bar{z} \quad (1.22)$$

Using the method described above, parameter ϕ_1 of equation (14) was estimated as 0.74. Since this estimate had a small standard error of 0.058 and a significant t-test value of 12.84, the estimate was good. The constant C was then found using equation (22). The estimated mean value and standard deviation of the residual a_t are -0.0778 and 26.0, respectively. The hypothesized expression of equation (1.14) representing the series becomes

$$z_t = -17.3 + 0.739 z_{t-1} + a_t \quad (1.23)$$

The next step is the diagnostic check to see if (1.23) does indeed represent the process. Equation (1.23) is used without the a_t term to estimate a series which is subtracted point-by-point from the data to obtain a residual series (represented by (1.18)). This residual series is the set of a_t (random shocks). The most important characteristics of the residuals are that they are uncorrelated and normally distributed. To determine this we find the ACF of the residuals. This is shown in Figure 1.15. For all lags greater than zero the ACF points are insignificant (t values < 1.96 = dashed line). A chi-square goodness of fit test for normality of the residuals with five degrees of freedom gives 2.4 which is less than 11.1 (for a level of significance of 5%) which indicates the residuals have a normal distribution. We conclude that the model is good.

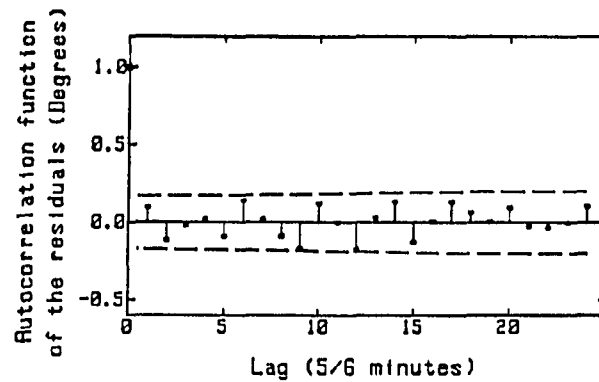


Figure 1.15. Autocorrelation of the residuals after a model was subtracted from the time series in Figure 1.2

We have modeled a wide variety of phase time series to see if the Box-Jenkins method can handle all of the various types of phase corruptions one would expect in the VLA data. We assume the equipment is working and have not considered things such as phase jumps caused by the local oscillator going out of lock. Several examples of modeling will be given illustrating the variety of data we have studied.

If the modeling process is satisfactory, we can test it further by predicting the phase behavior for future time. In the tests shown we have included the prediction at the end of the series but have omitted all intermediate details. Figures 1.16 and 1.17 show predictions at two different times for the same time series. In Figure 1.16 a model was built based on the first 80 min of data. The model is an

autoregressive moving average series with a first difference of the data:

$$\begin{aligned}
 z_t = & 0.7451 z_{t-1} + 0.5971 z_{t-2} - 0.1180 z_{t-3} - 0.2242 z_{t-4} \\
 & - 0.6318 z_{t-7} + 0.6318 z_{t-8} + 0.2729 a_{t-1} - 0.8568 a_{t-2} \\
 & - 0.6211 a_{t-3} - 0.5808 a_{t-7} + a_t
 \end{aligned} \tag{1.24}$$

Because the time series is nonstationary and the first difference was taken, there is no constant term C in the model. The model was used to make a pure prediction of the next 40 data points. The prediction is shown by the dashed line in Figure 1.16. A pure prediction of 40 points is far too long for a realistic prediction. However, by showing 40 points, the reader can see when the prediction starts to fail. In this case the prediction follows the actual data (shown by the solid line) reasonably well for about 9 points.

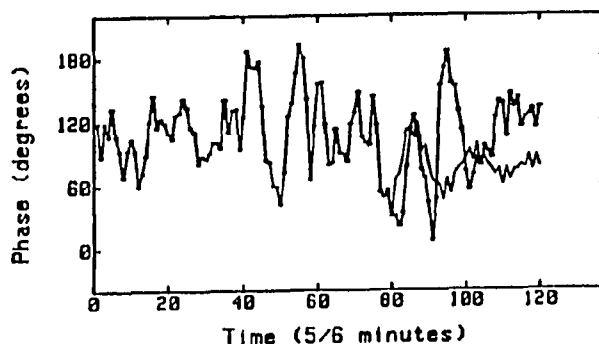


Figure 1.16. Forty-point pure prediction (dashed line) starting at point 80. The solid line is the data. The data were observed on April 7, 1984, with a wavelength of 2 cm on a 1.95-km baseline. The radio source was 3C 84

To illustrate another prediction with the same data, Figure 1.17 shows a pure prediction for 40 data points (dashed line) with a model based on the first 98 points. The prediction is good for about 10 points. The model for this prediction is different than (1.24) but contains the same number of terms. Of course, one can build a model with fewer terms, but in a case like this it may not represent the data well. This is shown by the prediction (dashed line) in Figure 1.18. The model is an AR(2). Comparing Figures 1.18 and 1.16 we see that the prediction in Figure 1.18 is inferior to that in Figure 1.16. The dash-dot lines in Figure 1.18 correspond to the 95% confidence interval for the prediction. Note that the prediction in Figure 1.18 quickly goes to the average, which is the best prediction when the known information is small.

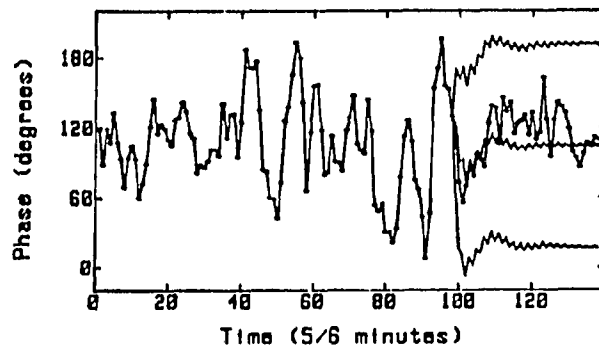


Figure 1.17. A pure prediction starting at point 98 (dashed line) of the same data as in Figure 1.16. The two dash-dot lines are the 95% confidence limits. The solid line is the data

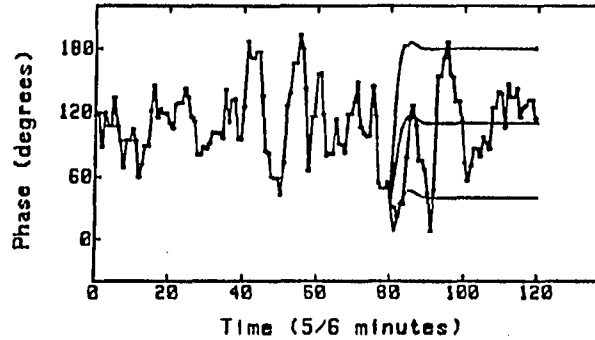


Figure 1.18. A pure prediction starting at point 80 using a simple AR(2) model. The line legend is the same as in Figure 1.17

Data of a more cyclical nature are shown in Figure 1.19. The model was built from the first 60 points and is

$$z_t = -2.041 + 1.6360 z_{t-1} - 0.8738 z_{t-2} - 0.5488 a_{t-1} + a_t \quad (1.25)$$

The model was used to predict the next 50 data points. The agreement with the actual data is good the first 17 points or so. The amplitude of the variation slowly decreases to the mean. The 95% confidence limits (dash-dot) lines encompass the actual data throughout the 50-point prediction. Another example of cyclical data is shown in Figure 20. We started the prediction at point 95 on a large swing of the data and continued for 50 points. The prediction follows the data remarkably well for about 20 points and then quickly converges to the mean. In

this case data do not deviate very much from the mean so the prediction is useful for a longer time period than we would normally expect.

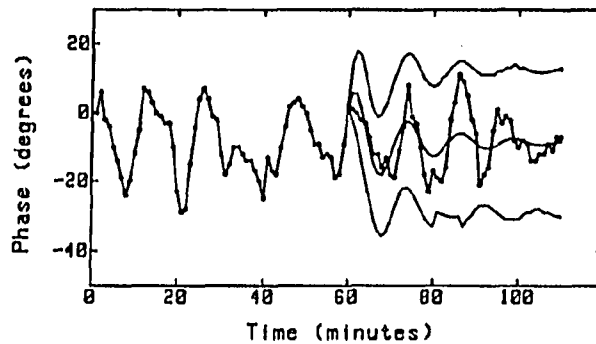


Figure 1.19. A 50-step pure prediction starting at point 60 of data with a cyclical behavior. The data were observed on March 7, 1982, at a wavelength of 6 cm. The radio source was 0316+161. The line legend is the same as in Figure 1.17

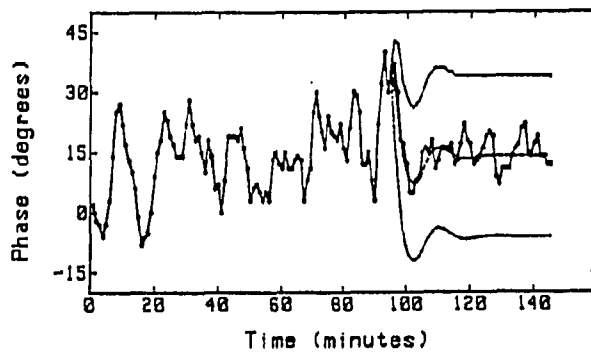


Figure 1.20. A 50-step pure prediction starting at point 95 of data observed on March 7, 1982, at a wavelength of 6 cm. The line legend is the same as in Figure 1.17

DISCUSSION

In the previous section, we discussed the procedure of finding stochastic models for VLA phase fluctuations. This procedure, often referred to as Box-Jenkins time series analysis, has four steps: (1) data preprocessing; (2) model identification; (3) parameter estimation; (4) model verification. The completeness of this procedure is necessary for obtaining a good model for a VLA phase time series.

Most of VLA phase time series are nonstationary. This nonstationarity shown in recorded data is due to the nonstationarity of random atmospheric variations. Phase data obtained with long geometric baselines show especially high nonstationarity. In this case, data differences or some nonlinear data transformations are needed to transform data series into approximately variance stationary series. The first difference was often taken for those data sampled at long baselines. Another useful and simple method is to segment the data series such that each segment is approximately variance stationary. From the results of our data analysis, we often had to create segments of about 50 to 70 min for long geometric baseline data. We found that the phase fluctuations of short baselines were well centrally distributed with small scatter and close to variance stationary. Consequently, the data differencing and segmentation were not often performed for the short-geometric-baseline data (such as baseline length < 2 km) even though the data were obtained with a high observing frequency and had large fluctuations. The variance stationarity of modeled data is very important for obtaining good predictions.

An ARIMA model of a VLA phase time series is data dependent. A variety of phase fluctuation types can be represented by a variety of ARIMA models based on intrinsic correlations existing in phase data. The results of modeling several VLA phase data sets have shown that most VLA phase series whose lengths are less than 120 minutes can be modeled by an ARIMA $(p,1,0)$ process where $p \leq 3$ or an ARIMA $(p,1,q)$ process where $p \leq 2$ and $q = 1$. A high order p of a model is helpful for longterm pure predictions of VLA phase fluctuations. Since MA terms quickly become negligible in a pure prediction procedure, a high order q is not recommended for constructing a model by which long-term predictions are derived. As the results show, a model with a high order of p produces better predictions than a model with a low order of p . However, the residuals generated from both models are very similar. Then, if we employ both small residual criteria and good predictions to judge a model, most VLA phase time series which we studied could be modeled by an ARIMA $(p,1,q)$ process where $p \leq 7$ and $q \leq 1$. This result is matched to the atmospheric variation which changes on a time scale of a few minutes if the sampling average varies from 30 s to 1 min.

In the estimation step for an ARIMA model, if the number of estimated parameters is small such as ≤ 8 , we often avoided solving the Yule-Walker equation and the Newton-Raphson algorithm for initial values of ϕ_i and θ_i . We assumed that all initial values were 0.1 and then applied Marquardt's method to obtain estimates of ϕ_i and θ_i . The

number of iterations for convergence often lies between 10 and 25. On the other hand, for an AR(p) process which requires all ϕ_i terms between ϕ_1 and ϕ_p , the method of solving the Yule-Walker equation was used to estimate ϕ_i . The results were very similar to those of Marquardt's method. When all intermediate coefficients are needed, solving the Yule-Walker equation is faster than Marquardt's method and easy to program.

The rms value of residuals generated by a model not only depends on the model, but also on the whiteness (uncorrelated) level of the phase data series. Those data obtained with either very short baselines or very long baselines with high observing frequencies usually have a high whiteness level. Then the rms values of residuals for these data modeled are relatively high. They may be close to, or over, 2/3 of the rms value of the data series. The rms value of residuals of 2-cm data are often relatively higher than those of 6-cm data. In the general case, if the phase variation is large, such as 30° rms, the residuals of the model can be reduced to about 15° to 20° rms, and good short term predictions can be made for 5 to 6 min before they deteriorate. For phase variations of around 10° rms, the residuals are reduced to about 4° to 5° rms. Pure predictions are typically good for 10 min or more.

Since our results show that ARIMA models of VLA phase fluctuations often provide good short-term predictions (some times for long-term predictions also), this modeling procedure may give good extrapolation or interpolation to help the empirical calibration.

The most important use of the models for us at this time is to put them into a Kalman filter. The time series modeling effectively separates the VLA phase variations in two components. One component is a nonwhite stochastic process represented by an ARIMA (p,d,q) model. The second component is white noise represented by the variance. The knowledge about these two components is necessary for applying the Kalman filter technique to improve radio images. An ARIMA model is a difference equation. Its Z transform, frequency response, and state-space form can be easily obtained. Then, many signal processing techniques can be used for further analysis and improvement. Currently, we are developing a technique for filtering data that is badly corrupted by atmospheric disturbances and testing the synthetic maps against the maps made by standard techniques.

ACKNOWLEDGMENTS

This project was funded by National Science Foundation grant no. AST- 8217135. We thank R. A. Sramek, T. J. Cornwell, and R. D. Ekers for helpful discussions throughout this project and thank A. A. Mitkees and F. M. M. El Mansy for many discussions at the initiation of this project. We thank Sramek for sharing VLA phase data. We thank the National Radio Astronomy Observatory for providing initial financial support for the project and for providing continued service support. The National Radio Astronomy Observatory is operated by Associated Universities, Inc., under contract with the National Science Foundation.

REFERENCES

- Anderson, O. D. 1976. Time series analysis and forecasting. Butterworth, Stoneham, Mass.
- Anderson, O. D., ed. 1982a. Time series analysis: theory and practice
1. Proc. of the International Conference held at Valencia, Spain, June 1981. North-Holland, New York, NY.
- Anderson, O. D., ed. 1982b. Time series analysis: theory and practice
2. Proc. of the International Conference held at Dublin, Ireland, March 1982. North-Holland, New York, NY.
- Anderson, O. D., ed. 1983. Time series analysis: theory and practice
4. Proc. of the International Conference held at Cincinnati, Ohio, August 1982. North-Holland, New York, NY.
- Armstrong, J. W., and R. A. Sramek. 1982. Observations of tropospheric phase scintillations at 5 GHz on vertical paths. Radio Science 17:1579-1586.
- Baars, J. W. M. 1967. Meteorological influences on radio interferometer phase fluctuations. IEEE Trans. Antennas Propagation AP-15:582-584.
- Box, G. E. P., and G. M. Jenkins. 1976. Time series analysis: forecasting and control. 2nd ed. Holden-Day, San Francisco, CA.
- Brown, R. G. 1983. Introduction to random signal analysis and Kalman filtering. John Wiley & Sons, New York.
- Clark, B. G. 1980. An efficient implementation of the algorithm "CLEAN". Astron. Astrophys. 89:377-378.
- Cornwell, T. J. 1983. A method of stabilizing the clean algorithm. Astron. Astrophys. 121:281-285.
- Dravskikh, A. F., and A. M. Finkelstein. 1979. Tropospheric limitations in phase and frequency coordinate measurements in astronomy. Astrophys. and Space Sci. 60:251-265.
- Hamaker, J. P. 1978. Atmospheric delay fluctuations with scale sizes greater than one kilometer, observed with a radio interferometer array. Radio Sci. 13:873-891.
- Han, W. -J. 1980. Effects of atmosphere irregularities on radio interferometer measurements. Chinese Astronomy 4:174-184.

- Hargrave, P. J.; and L. J. Shaw. 1978. Large-scale tropospheric irregularities and their effect on radio astronomical seeing. *Mon. Notic. Royal Astron. Soc.* 182:233-239.
- Hinder, R. A. 1970. Observations of atmospheric turbulence with a radio telescope at 5 GHz. *Nature* 225:229-253.
- Hinder, R., and M. Ryle. 1971. Atmospheric limitations to the angular resolution of aperture synthesis radio telescopes. *Mon. Notic. Royal Astron. Soc.* 154:229-253.
- Hogbom, J. A. 1974. Aperture synthesis with a non-regular distribution of interferometer baselines. *Astron. Astrophys. Suppl.* 15:417-426.
- Hogbom, J. A. 1984. CLEAN as a pattern recognition procedure. Pages 247-254 *in* J. A. Roberts, ed. *Indirect Imaging*. Cambridge University Press, Cambridge, NY.
- Hogg, D. C., F. O. Guiraud, and W. B. Sweezy. 1981. Measurement temporal spectrum of precipitable water vapor. *Science* 213:1112-1113.
- Ishimaru, A. 1978. Wave propagation and scattering in random media. Vol. 2. Academic, Orlando, Fla.
- Jennison, R. C. 1958. A phase sensitive interferometer technique for the measurement of the Fourier transforms of spatial brightness distributions of small angular extent. *Mon. Notic. Royal Astron. Soc.* 118:276-284.
- Kalman, R. E. 1960. A new approach to linear filter and prediction problems. *Trans. ASME* 82:35-45.
- Kalman, R. E., and R. S. Bucy. 1961. New results in linear filtering and prediction theory. *Trans. ASME* 83:95-108.
- Marquardt, D. W. 1963. An algorithm for least squares estimation of non-linear parameters. *J. Soc. Ind. Appl. Math.* 11:431-441.
- Mathur, N. C., M. D. Grossi, and M. R. Pearlman. 1970. Atmospheric effects in very long baseline interferometry. *Radio Sci.* 5:1253-1261.
- McCleary, R., and R. A. Hay, Jr. 1980. *Applied series analysis for the social sciences*. Sage, Beverly Hills, Calif.
- Moran, J. M., and B. R. Rosen. 1981. Estimation of the propagation delay through the troposphere from microwave radiometer data. *Radio Sci.* 16:235-244.

- Napier, P. J., A. R. Thompson, and R. D. Ekers. 1983. The very large array: design and performance of a modern synthesis radio telescope. *Proc. IEEE* 71:1295-1320.
- Okatan, A., and J. P. Basart. 1979. Reduction of phase noise in interferometry with state-space analysis. Pages 103-110 in C. Van Schooneveld, ed. *Image formation from coherence functions in astronomy*. Vol. 76. D. Reidel, Hingham, Mass.
- Pankratz, A. 1983. *Forecasting with univariate Box-Jenkins models*. John Wiley, New York, NY.
- Readhead, A. C. S., and P. N. Wilkinson. 1978. The mapping of compact radio sources from VLBI data. *Astrophys. J.* 223:25-36.
- Resch, G. M., D. E. Hogg, and P. J. Napier. 1984. Radiometric correction of atmospheric path length fluctuations in interferometric experiments. *Radio Sci.* 19:411-422.
- Rogers, A. E. E., et al. 1974. The structure of radio sources 3C 273b and 3C 84 deduced from the "closure" phases and visibility amplitudes observed with three-element interferometers. *Astrophys. J.* 193:293-301.
- Schwab, F. R. 1980. Adaptive calibration of radio interferometer data. *Proc. Soc. Photogr. Optic. Instrum. Eng.* 231:18-25.
- Schwarz, U. T. 1978. Mathematical-statistical description of the iterative beam removing technique (method CLEAN). *Astron. Astrophys.* 65:345-356.
- Thompson, A. R., B. G. Clark, C. M. Wade, and P. J. Napier. 1980. The very large array. *Astrophys. J. Suppl.* 44:151-167.
- Wesseling, K. H., J. P. Basart, and J. L. Nance. 1974. Simultaneous interferometer phase and water vapor measurements. *Radio Sci.* 9:349-353.
- Westwater, E. R., and F. O. Guiraud. 1980. Ground-based microwave radiometric retrieval of precipitable water vapor in the presence of clouds with high liquid content. *Radio Sci.* 15:947-957.

PART II. IMAGE ANALYSIS, FEATURE EXTRACTION, AND VARIOUS APPLIED
ENHANCEMENT METHODS FOR NDE X-RAY IMAGES

ABSTRACT

This work introduces the main features of a relatively complete procedure for NDE X-ray image processing. The procedures include image analysis, feature extraction, and image enhancement. Image analysis uses information of image statistics and image contents. Image feature extraction includes an image transformation, an image segmentation using a split and merge method, and edge detection by the masking function. Image enhancement techniques are evaluated for a low pass filter, sigma filter, median filter, adaptive variance filter, and the Kalman filter for different application situations. Many examples are given to illustrate different techniques.

INTRODUCTION

As one of the "big five" NDE methods, radiography has been widely applied in industrial inspection procedures to detect flaws with a wide range of attainable flaw sensitivities [Halmshaw, 1982]. As one of the radiographic methods, X-ray radiography for NDE has been steadily developed for more than 40 years and is now well established. However, quality of an X-ray image is limited by intrinsic characteristics of a radiographic image forming mechanism [Halmshaw, 1982; Halmshaw, 1973; Herz, 1969]. The limitations cause difficulties in defining a fine structure of a specimen and to detect a weak flaw in a low contrast radiograph.

The objective of this work is to introduce a general procedure to enhance a low-contrast radiographic X-ray image and evaluate flaws in the image. This procedure includes image analysis, image enhancement, and image feature extraction. In this procedure, contents of a radiographic image are analyzed to determine a proper method for enhancing this specific image, and a feature extraction method is chosen to detect flaws.

Image analysis focuses on three factors: image contrast, unsharpness, and film graininess. Various enhancement techniques proposed to improve these factors include the median filter, sigma filter, adaptive smoothing filter, Kalman filter, histogram equalization, and background trend removal. Although blurring is mentioned here, its solution is not included at this time. Feature extraction methods applied include

flaw spatial activity detection by the masking function, and flaw detection by image segmentation.

X-RAY RADIOGRAPHIC IMAGE ANALYSIS

X-rays are electromagnetic radiations with wavelengths roughly in the range of 10^{-10} to 10^{-7} cm which are shorter than those of visible light and ultraviolet. X-rays are generated by accelerating a high voltage beam of electrons onto a metal target. Because of their penetrating property, X-rays are widely used in industrial inspection. Three fundamental parameters of a radiographic image: contrast, unsharpness, and film graininess [Halmshaw, 1982] are briefly analyzed with methods proposed to improve each of them.

Image Contrast

Image contrast is defined by the density difference between the image of a slit and the background given by a uniformly thick specimen. Whether the slit is discernible depends on the amount of the difference. The shallower the discernible slit is, the higher the image contrast must be. Hence the contrast is related to the ability to detect a minimum thickness change in a specimen. Assuming that the smallest observable thickness change is Δx in a specimen of thickness x , the contrast sensitivity is defined as [Halmshaw, 1982]

$$\frac{\Delta x}{x} \cdot 100 = \frac{2.3 \Delta D [1 + (I_s/I_D)]}{\mu G_D x} \cdot 100\% \quad (2.1)$$

where ΔD = film density change due to the thickness change Δx , I_s = intensity due to scattered radiation, I_D = intensity of the direct

radiation, μ = linear absorption coefficient of the exponential absorption [Halmshaw, 1982], and G_D slope constant of the linear region of a film characteristic curve $D\text{-log}E$ [Halmshaw, 1982]. To obtain a high contrast image, high-contrast film having a high sensitivity and fine graininess should be used, and an X-ray radiation with a proper energy should be adjusted.

For a specimen with a fixed x , the attainable thickness sensitivity can be improved by reducing the effect of I_s/I_D in equation (2.1) by removing the scattered radiation. If the scattering effect appears with an additional uniform intensity, it can be reduced by removing the background and properly expanding the image's dynamic range. If the scattering effect causes a random fluctuation, it can be reduced by a proper filtering procedure. ΔD has a value between 0.006 and 0.01 Depending on viewing conditions [Halmshaw, 1982]. In the denominator of equation (2.1), x is fixed for a specific specimen and μ is related to the material and the absorption process. To obtain a high contrast image, strong absorption generated by low-energy radiation is needed. However, the radiated energy should match the thickness of a specimen. Therefore, with a specific material, μ is controlled by a compromise between the contrast sensitivity and the penetrating power. Another term in the denominator, G_D is determined by the film contrast and the film exposure. If a film is over exposed or under exposed, G_D is roughly given by the slopes at the two ends of a film characteristic curve, which is very small. With a small G_D , the

value of Δx is high and the image contrast is very low. If the film is under exposed, the image should be reconstructed with a reconstruction method for missing information. If it is over exposed, a trend removing method may be used to remove the background intensity, and then, a dynamic range expanding method or histogram equalization used to obtain an improved image with increased contrast.

Another factor not considered in equation (2.1), is contrast reduction caused by quantum fluctuation due to random emission and absorption of X-ray quanta. With known statistics of the fluctuation, a filter can be chosen to reduce this effect.

Unsharpness

Unsharpness refers to the blurred image of a sharp edge of a specimen. There are at least two main causes of radiographic unsharpness: geometric unsharpness and film unsharpness [Halmshaw, 1982]. Geometric unsharpness is caused by the finite size of the radiating source. It is exactly analogous to a penumbra in visible light. This unsharpness can be avoided by using a point source, or alternatively, increasing the distance between the radiation source and the specimen. However, the intensity of the radiation decreases with distance by an inverse square law, and thus, rapidly increases the required exposure time. A compromise has to be made between geometric unsharpness and the distance.

Film unsharpness is caused by the scattering of electrons in the film emulsion. With high energy radiation, several hundred silver halide crystals may be affected by the absorption of one X-ray quantum in the film emulsion which may mask an image of a sharp edge.

Assuming that the intensity of an image is represented by $x(i,j)$, unsharpness can be described by

$$y(i,j) = \sum_{k=i-M}^{i+M} \sum_{l=j-N}^{j+N} x(k,l)b(k,l) \quad (2.2)$$

where $b(k,l)$, a point-spread function describing the blurring, has a size determined by the width of the unsharpness. To obtain $x(i,j)$ from observations $y(i,j)$, a deblurring method should be developed. The critical condition is that the point-spread function $b(i,j)$ must be known.

Film Graininess

Film graininess, a granular noise appearance of an image, is caused by a natural statistical fluctuation of randomly distributed silver particles. Many efforts have been made to find its statistics. In Nutting [1913] expressed the image density in terms of the number and size of grains. Based on this model, Selwyn [1939; 1959], related the standard deviation $\sigma_{\Delta D}$ of the film-grain noise to the scanning spot size, A , by

$$\sigma_{\Delta D} = G A^{-1/2} \quad (2.3)$$

where G is the Selwyn granularity constant. Selwyn's formula is true only in certain circumstances. In Johns [1955] introduced a spectral analysis method to obtain $\sigma_{\Delta D}$ by

$$\sigma_{\Delta D} = (\int W(f)df)^{1/2} \quad (2.4)$$

where $W(f)$ is the Wiener spectrum of film-grain noise. If the spectral components of the specimen are known or have certain patterns, there is a definite advantage to using the Wiener spectrum to study film-grain noise. In several practical applications in the last two decades, a model relating $\sigma_{\Delta D}$ to the optical density was used for image estimations and enhancements [Huang, 1966; Naderi and Sawchuk, 1978; Kuan et al., 1985]. This model is

$$\sigma_{\Delta D} = kD^{1/3} \quad (2.5)$$

where k is a constant related to the ratio of the grain size to the scanning spot size, and D is the optical density of the image. Equation (2.5) shows that the noise is signal dependent.

Film-grain noise smears the fine detail and reduces the contrast of an image. If we define the signal-to-noise ratio by the ratio of signal intensity $(D-M_b)$ to standard deviation $\sigma_{\Delta D}$ where M_b is a mean of the background intensity, a signal-to-noise ratio of above 2 is required to detect a signal, and a signal-to-noise ratio of above 4 is required to recognize a signal. A simple and direct method to reduce $\sigma_{\Delta D}$ is to increase the size of A [Dainty and Shaw, 1974]. However,

increasing A will decrease the resolution of a digitized image. If we know the characteristics of the film-grain noise, we can choose a proper filter to reduce it.

Many articles and text books [Herz, 1969; Dainty and Shaw, 1974] show that film-grain noise is spatially clustered and correlated. In addition to demonstrations given by micrographs of film-grain spatial structures, the correlation can be readily found from the Wiener spectrum of the film-grain noise which mainly consists of low-frequency components. The level and flatness of the spectrum depend on the quality of a film [Halmshaw, 1982]. Very-fine-grain film has a relative low and flat spectrum distribution approximately modeled by white noise. Otherwise, for a general case, film-grain noise should be modeled by colored (correlated) noise whose correlation distance is related to the scanning spot size, A , and the X-ray energy used. In addition, the blurring process of chemical diffusion and adjacency effects during the film development will increase the correlation distance. Hence, noise appears to be spatially correlated in an observed image. Our calculation of the autocorrelation coefficients for noise fluctuations verifies that the noise is spatially correlated. The autocorrelation coefficients, found in the flat areas of the images, vary at lag 1 from 0.2 to 0.5 for the images displayed later. The correlation distance varies from 2 to 4 pixels.

The probability distribution of optical density has been reported to be nearly Gaussian for a particular dynamic range of a film [James,

1977]. A Gaussian distribution is also assumed for the noise fluctuation [Dainty and Shaw, 1974], but, this is not necessarily true for every image met in practice.

Based on results of the image analysis, a possible approach can be chosen for image enhancement. If the film is processed in the linear region of the D-logE curve, the film grain noise is additive in an observed image. Thus, considering the above discussions, a specific filter approach is chosen to reduce the noise level for a specific image. We suggest: reduce salt and pepper noise with a median filter, reduce Gaussian white noise with a sigma filter, reduce signal-dependent white noise with an adaptive smoothing filter, reduce colored noise or signal dependent colored noise with a Kalman filter, reduce the effect of over exposure with a trend removal and a dynamic range expansion or histogram equalization, improve an under exposed image with image reconstruction by considering missing information, and finally, improve sharpness of an image with a deblurring algorithm.

ENHANCEMENT METHODS

Various applied enhancement techniques, except deblurring and reconstruction, are briefly described in this section. To obtain good results, try several of them iteratively on an image. A combination of several methods can often produce good results.

Trend Removal

Trend removal removes the large scale intensity variation caused by either the geometry of a specimen or over exposure. Combining this method with a dynamic range expansion or histogram equalization can reveal masked fine structures or flaws. This work is well discussed and documented by Doering [Doering, 1987]. The basic idea of a trend removal method is to fit a polynomial model to the background trend by least-squares, and then find the residuals between the image and the polynomial model. This fit could be for an entire image or for a running window. It could also be either one dimensional or two dimensional. As an example, we fit a second-order one-dimensional polynomial to samples y_i by

$$P_i = a + bi + ci^2 \quad (2.6)$$

where i is the location of a sample and coefficients a , b , c are found by minimizing the summation of the squared error

$$Q = \sum_{i=1}^N (y_i - p_i)^2 \quad (2.7)$$

where N is a window size. Taking a derivative with respect to each coefficient of equation (2.7) and equating to zero, three linear equations can be formed in three unknown coefficients. Subsequently, the three coefficients can be determined.

Dynamic Range Expansion

If an image has a small dynamic range, the structure of the image may not be easily discerned by eye. In this case, a simple formula is useful. The small dynamic range is expanded by

$$Y_i = Y_{\min} + (y_i - y_{\min})(Y_{\max} - Y_{\min}) / (y_{\max} - y_{\min}) \quad (2.8)$$

where y_i is an intensity value of the image and Y_{\max} and Y_{\min} are maximum and minimum values in the expanded image, respectively. y_{\max} and y_{\min} are maximum and minimum values in the original image. The expansion ratio is given by $(Y_{\max} - Y_{\min}) / (y_{\max} - y_{\min})$. The value of $(Y_{\max} - Y_{\min})$ depends on the capability of a specific image display system. Equation (2.8) is linear, but a nonlinear expansion may be also helpful.

Histogram Equalization

Histogram equalization is useful for enhancing an image whose gray levels are in a narrow range and a few pixels are sparsely distributed over a large dynamic range. If the interesting information is in the narrow range, it can be revealed by modifying the histogram of gray levels to a histogram with a uniformly distributed probability density function. Let a sample be normalized by y_i/y_{\max} the new value of this sample after the histogram equalization is given by [Gonzalez and Wintz, 1977].

$$Y_i = \int_0^{y_i} p_i(u) du \quad (2.9)$$

where $p_i(u)$ is a probability density function for the gray level y_i . p_i can be readily obtained from the histogram. The enhanced image can be obtained by magnifying Y_i with equation (2.8).

Median Filter

A median filter is a nonlinear filter used to reduce impulse (or salt and pepper) noise with the advantage of preserving edge information. There are many forms of median filters. Two recent papers [Arce and McLoughlin, 1987; McLoughlin and Arce, 1987] give good reviews of this technique and report the latest progress. A two-dimensional $N \times N$ recursive median filter is given by

$$Y_{i,j} = \text{median} \{Y_{i-N,j}, \dots, Y_{i-1,j}, z_{i,j}, \dots, z_{i+N,j}\} \quad (2.10)$$

where

$$z_{m,n} = \text{median} \{z_{m,n-N}, \dots, z_{m,n-1}, y_{i,j}, \dots, y_{m,n+N}\}$$

and $y_{i,j}$ is an estimation of the median filter for sample $y_{i,j}$.

Sigma Filter

The sigma filter is designed for smoothing Gaussian white noise. It averages pixels whose gray levels are in the $\pm 2\sigma_n$ range for a given pixel within a window [Lee, 1984] where σ_n is the standard deviation of the noise. A $(2N+1) \times (2N+1)$ two-dimensional sigma filter is given by

$$Y_{ij} = \begin{cases} \sum_{k=i-N}^{i+N} \sum_{l=j-N}^{j+N} y_{kl} w_{kl} / \text{NUM}_{ij}, & \text{if } \text{NUM}_{ij} > K \\ \sum_{k=i-N}^{i+N} \sum_{l=j-N}^{j+N} y_{kl} / (2N+1)^2, & \text{if } \text{NUM}_{ij} < K \end{cases} \quad (2.11)$$

where

$$w_{kl} = \begin{cases} 1, & \text{if } y_{ij} - 2\sigma_n < y_{kl} < y_{ij} + 2\sigma_n \\ 0, & \text{otherwise} \end{cases}$$

and

$$\text{NUM}_{ij} = \sum_{k=i-N}^{i+N} \sum_{l=j-N}^{j+N} w_{kl}$$

The value of K is chosen so that isolated spot noise can be removed without destroying subtle details of an image. The sigma filter may be used iteratively.

Adaptive Smoothing Filter

The adaptive smoothing filter is designed to smooth signal dependent white noise. There are several forms of adaptive smoothing filters [Lee, 1980; Kuan et al., 1985]. Assuming that an observation equation is given by

$$y_{ij} = x_{ij} + u_{ij} \quad (2.12)$$

where u_{ij} is an $iid(0, V_u(x_{ij}))$ and $V_u(x_{ij})$ is a semi-positive-definite function of the signal x_{ij} . u_{ij} can be represented by $f(x_{ij})n_{ij}$ where $f(x_{ij})$ means signal dependent and n_{ij} is an $iid(0,1)$. The principle of an adaptive smoothing filter is to estimate x_{ij} by first predicting a local mean and then correcting this by a calibration factor. The estimation of x_{ij} is given by

$$\hat{x}_{ij} = \bar{x}_{ij} + k_{ij}(y_{ij} - \bar{x}_{ij}) \quad (2.13)$$

where \bar{x}_{ij} is a local mean of x_{ij} equaling \bar{y}_{ij} because $E[u_{ij}] = 0$. \bar{x}_{ij} is calculated from an $N \times N$ local window. k_{ij} , a calibration gain found by minimizing the least-squared error, is given by

$$k_{ij} = V_{x_{ij}} / (V_{x_{ij}} + V_u(x_{ij})) \quad (2.14)$$

where $V_{x_{ij}}$ is a variance of the signal x_{ij} and $V_u(x_{ij})$ is a variance of the noise u_{ij} . If the signal-to-noise ratio is very high, k_{ij} is close to one and the estimate is close to the measurement. If the signal-to-noise ratio is low, k_{ij} is close to zero and the estimate is close to

the local mean. Since $E[n_{ij}] = 0$, the covariance between x_{ij} and u_{ij} is zero and $V_{y_{ij}} = V_{x_{ij}} + V_{u_{ij}}$ where $V_{y_{ij}}$ is a variance of y_{ij} . Thus, k_{ij} can be calculated by $(V_{y_{ij}} - V_u(x_{ij})) / V_{y_{ij}}$. This filter is similar to a one-dimensional Kalman filter with a predicted state given by a mean value.

Kalman Filter for Image Restoration

A Kalman filter recursively separates two or more stochastic processes with the criterion of least-squared error. The spectra of those processes may overlap preventing classical filters such as low-pass or high-pass filters from separating the processes. The theory of the Kalman filter is introduced in many papers and text books [Kalman, 1960; Kalman and Bucy, 1961; Gelb, 1974; Brown, 1983]. The Kalman filter was first introduced in the control area and has been widely applied to control systems, navigation systems, and random signal processing systems. The application of Kalman filtering to image restoration was intensively investigated in the last decade [Woods and Radewan, 1977; Suresh and Shenoi, 1981; Rajala and de Figueiredo, 1981; Dikshit, 1982; Biemond et al., 1983; Tekalp et al., 1986].

As a simple example, let a vector of observations be given by

$$y_{ij} = x_{ij} + v_{ij} + n_{ij} \quad (2.15)$$

where x_{ij} is a signal vector to be estimated, v_{ij} is a colored noise vector, and n_{ij} is white noise vector. The estimation vector, \hat{x}_{ij} , given by a Kalman filter is

$$\hat{x}_{ij} = \hat{x}_{ij}^- + K_{ij} (y_{ij} - \hat{x}_{ij}^- - \hat{v}_{ij}^-) \quad (2.16)$$

where $\hat{\mathbf{x}}_{ij}^-$ is the a priori estimate of $\hat{\mathbf{x}}_{ij}$ predicted by a signal model, and $\hat{\mathbf{v}}_{ij}^-$ is the a priori estimate of \mathbf{v}_{ij} given by a noise model. \mathbf{K}_{ij} is an optimal gain vector. More details of the application of the Kalman filter to radiographic image enhancement are given by Zheng and Basart [1988].

IMAGE FEATURE EXTRACTION

Image feature extraction is a process of image classification. Image features refer to meaningful image components and their inter relation. They are classified by intensity, contrast, edge, size, and shape of spatial activity. Detected features may be structures of a specimen, or significant flaws or cracks. Three useful methods discussed here are spatial activity detection, image segmentation, and significance testing.

Flaw Spatial Activity Detection

Spatial activity of an image is related to the concept of large scale spatial fluctuations measured by variances and covariances, or the concept of small scale spatial variation measured by gradients. We use a modified masking function to identify spatial activity of an image. The masking function at a pixel is a weighted sum of the intensity slopes at the pixel under consideration and at the neighboring pixels [Rajala and de Figueiredo, 1981]. The masking function also plays a role as an edge detector. There are many edge detection methods and most of them are sensitive to noise. This modified masking function can detect spatial activity of moderately sized flaws in a low signal-to-noise ratio image where a conventional edge detection method may fail. The masking function is defined as

$$M_r(i, j) = \frac{1}{C} \sum_{p=i}^{i+r} \sum_{q=j-r}^{j+r} a ||(i, j), (p, q)|| \sum_{k=1}^4 D_{pqk} \quad (2.17)$$

where a is a constant less than one, r is a window size, $||(i, j) - (p, q)||$ is the Euclidean distance, and D_{pqk} is a slope of a pixel (p, q) in a direction of k . The value of D_{pqk} can evaluated by

$$D_{pqk} = \left| \sum_{i=1}^M y_{(p+ai)(q+bi)} - \sum_{i=1}^M y_{(p+ci)(q+di)} \right| \quad (2.18)$$

where

$$\begin{aligned} a = 0, b = 1, c = 0, d = -1; & \quad \text{if } k = 1 \\ a = 1, b = 1, c = -1, d = -1; & \quad \text{if } k = 2 \\ a = 1, b = 0, c = -1, d = 0; & \quad \text{if } k = 3 \\ a = 1, b = -1, c = -1, d = 1, & \quad \text{if } k = 4, \end{aligned}$$

and $|\cdot|$ indicates absolute value, and M is a smoothing length. The smoothing length M should be larger than the significant correlation distance of the noise and smaller than the size of the fine structure to be detected. Since D_{ijk} is formed in a different way than that of the standard masking function [Anderson and Netravali, 1976; Netravali and Brasada, 1977], the function given by equation (17) and (18) is called the modified masking function. We have applied the modified masking function to X-ray radiographic images. Figure 2.1a shows a digitized radiograph of a fuel-tank weld from the Martin Marietta

Corporation. The image size, 256 x 256, represents a physical area of $8.2 \times 8.2 \text{ mm}^2$. The trend of the image was removed by the method mentioned above. If the SNR is defined by the ratio between the relative intensity of the flaw and the standard deviation of the noise [Zheng and Basart, 1988], the signal-to-noise ratio of Figure 1.1a is about 5.0 for the strongest flaw. A Kalman filter was first applied to the image to reduce the noise level [Zheng and Basart, 1988] followed by a sigma filter to further enhance the image. The modified masking function was then used to detect the spatial activity of the filtered image with the result given by Figure 2.1b. One can see that the spatial activity of the flaws is well detected while the effect of the noise is small. However, the outlines of flaws are not closed lines and are wider than that given by a conventional gradient method or a conventional masking function. This is a price paid for reducing the noise effect.

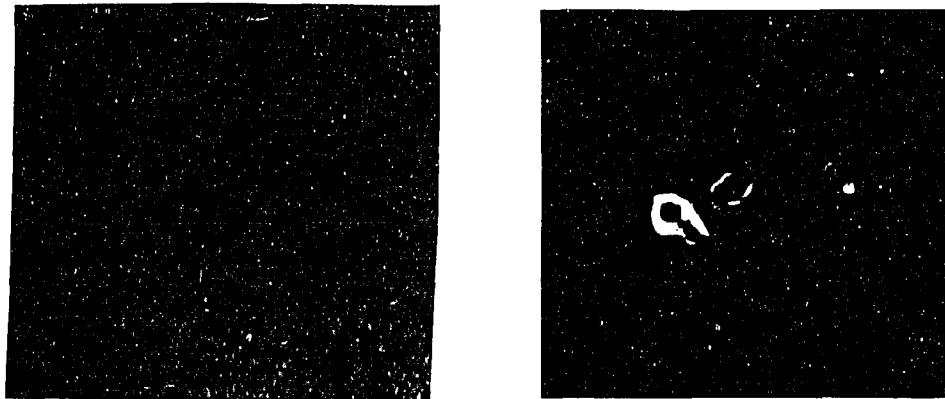


Figure 2.1. (a) A digitized NDE radiographic X-ray image of a weld from the Martin Marietta Corporation, (b) an output of the masking function for a filtered image of (a).

Flaw Detection by Image Segmentation

There are two types of image segmentations: low-level and high-level. Low-level image segmentation partitions an image according to statistical numbers while high-level image segmentation partitions an image according to its natural physical contents [Gonzalez and Wintz, 1977; Schachter et al., 1979; Nazif and Levine, 1984; Stansfield, 1986]. There are many image segmentation methods such as optimal thresholding [Gonzalez and Wintz, 1977], split-and-merge methods [Chen and Pavlidis, 1979], model-based methods [Chatterjee and Chellappa, 1987], the Gibbs distribution method [Elliott et al., 1986], region growing and region clustering [Gonzalez and Wintz, 1977], K-mean method and isodata method [Tou and Gonzalez, 1974]. Fuzzy c-mean clustering [Cannon et al., 1986], and rule-based expert systems [Nazif and Levine, 1984; Stansfield, 1986; Zheng and Basart, 1987].

We have used a low-level image segmentation method and a rule-based system to partition an image [Zheng and Basart, 1987]. We have found flows with simple thresholding based on a significance test. The darkness of a pixel, which is related to density of a film, is significant if it is less than a threshold

$$d_{ij} = \begin{cases} 1; & \text{if } y_{ij} < \text{mean of the background noise} - K\sigma_n \\ 0; & \text{otherwise} \end{cases} \quad (2.19)$$

where σ_n is the standard deviation of the noise and $K > 2.0$. Then, a binary file is formed by d_{ij} . This procedure was used to detect flaws

in the filtered image of Figure 2.1a with the result shown in Figure 2.2. Three flaws are easily seen in Figure 2.2, whereas in Figure 2.1 it is difficult to see the right flaw. The shape and size are also well displayed in Figure 2.2.

The techniques described herein have been implemented on the Iowa State University VAX 11/780 computers. Among the above steps discussed, we feel that image analysis is a critical step. Conclusions about the image analysis will be main reasons for choosing specific enhancement methods. In addition, the accuracy of conclusions or models obtained from the image analysis will influence the results of the enhancement and feature extraction no matter what filters are used.



Figure 2.2. Plot of detected flaws of Figure 2.1(a) after the processing

ACKNOWLEDGMENTS

We thank James Lloyd and David Fischer of Martin Marietta for providing radiographs, and thank Joseph Gray for many helpful discussions. This work is supported by the NSF University/Industry Center for NDE at Iowa State University.

REFERENCES

- Anderson, G. L., and A. N. Netravali. 1976. Image restoration based on a subjective criterion. *IEEE Trans. Syst. Man Cyb.* CMS-6:845-853.
- Arce, G. R., and M. P. McLoughlin. 1987. Theoretical analysis of the max/median filter. *IEEE Trans. Acoust. Speech Signal Processing ASSP-35*:60-69.
- Biemond, J., J. Rieske, and J. J. Gerbrands. 1983. A fast Kalman filter for images degraded by both blur and noise. *IEEE Trans. Acoust., Speech, Signal Processing ASSP-31*:1248-1256.
- Brown, R. G. 1983. Introduction to random signal analysis and Kalman filtering. John Wiley & Sons, New York.
- Cannon, R. L., J. V. Dave, and J. C. Bezdek. 1986. Efficient implementation of the fuzzy c-means clustering algorithms. *IEEE Trans. Pat. Anal. Mach. Intel. PAMI-8*:248-255.
- Chatterjee, S., and R. Chellappa. 1987. Model-based texture segmentation. Pages 235-239 in D. M. Etter, ed. Twentieth Asilomar Conference on Signal, Systems & Computers. IEEE Inc., New York.
- Chen, P. C., and T. Pavlidis. 1979. Segmentation by texture using a co-occurrence matrix and a split-and-merge algorithm. *Comp. Graph. Image Proc.* 10:172-182.
- Dainty, J. C., and R. Shaw. 1974.. Image science. Academic Press Inc., New York.
- Doering, E. R. 1987. Detection of anomalies in digital images using pixel classification. M.S. Thesis. Iowa State University, Ames, Iowa.
- Dikshit, S. S. 1982. A recursive Kalman window approach to image restoration. *IEEE Trans. Acoust., Speech, Signal Processing ASSP-30*:125-140.
- Elliott, H., H. Derin, and R. Cristi. 1986. Application of the Gibbs distribution to image segmentation. Pages 3-24 in E. J. Wegman and D. J. Depriest, ed. *Statistical Image Processing and Graphics*. Marcel Dekker, Inc., New York.
- Gelb, A., ed. 1974. Applied optimal estimation. MIT Press, Cambridge, Mass.

- Gonzalez, R. C., and P. Wintz. 1977. Digital Image Processing. Addison-Wesley, Reading, Mass.
- Halmshaw, R. 1982. Industrial radiology: theory and practice. Applied Science Publishers Ltd. Englewood, New Jersey.
- Halmshaw, R. 1973. Information theory applied to industrial radiography. Pages 31-64 in R. S. Sharpe, ed. Research techniques in nondestructive testing. Vol. 2. Academic Press, New York.
- Herz, R. H. 1969. The photographic action of ionizing radiations. John Wiley & Sons Inc, New York, NY.
- Huang, T. S. 1966. Some notes on film grain noise. Pages 105-109 in Woods Hole Summer Study Report. Restoration of atmospherically degraded images. Vol. 2, Appendix 14. National Academy of Sciences and National Research Council, Washington DC.
- James, T. H. 1977. The theory of the photographic process. Macmillan, New York.
- Johns, R. C. 1955. New method of describing and measuring the granularity of photographic materials. J. Opt. Soc. America 45:799-808.
- Kalman, R. E. 1960. A new approach to linear filter and prediction problems. Trans. ASME 82:35-45.
- Kalman, R. E., and R. S. Bucy. 1961. New results in linear filtering and prediction theory. Trans. ASME 83:95-108.
- Kuan, D. T., A. A. Sawchuk, T. C. Strand, and P. Chavel. 1985. Adaptive noise smoothing filter for images with signal-dependent noise. IEEE Trans. Pattern Anal. Mach. Intel. 7:165-177.
- Lee, J. S. 1984. The sigma filter and its application to speckle smoothing of synthetic aperture radar images. Pages 445-459 in J. W. Eward and G. S. Smith ed. Statistical Signal Processing. Marcel Dekker, New York, NY.
- Lee, J. S. 1980. Digital image enhancement and noise filtering by use of local statistics. IEEE Trans. Pattern. Anal. Mach. Intel. PAMI-2:165-168.
- McLoughlin, M. P., and G. R. Arce. 1987. Deterministic Properties of the recursive separable median filter. IEEE Trans. Acoust. Speech Signal Processing ASSP-35:98-106.
- Naderi, F., and A. A. Sawchuk. 1978. Estimation of Images Degraded by Film-Grain Noise. Applied Optics 17:1228-1237.

- Nazif, A. M. and M. D. Levine. 1984. Low level image segmentation: An expert system. *IEEE Trans. Pattern Anal. Machine Intell.* PAMI-6:555-577.
- Netravali, A. N., and B. Brasada. 1977. Adaptive quantization of picture signals using spatial masking. *Proc. IEEE* 65:536-548.
- Nutting, P. G. 1913. On the absorption of light in heterogeneous media. *Philos. Mag.* 26:423-426.
- Rajala, S. A., and de Figueiredo, R. I. P. 1981. Adaptive nonlinear image restoration by a modified Kalman filtering approach. *IEEE Trans. Acoust., Speech, Signal Processing* ASSP-29:1033-1042.
- Schachter, B. J., L. S. Davis, and A. Rosenfeld. 1979. Some experiments in image segmentation by clustering of local feature values. *Pattern Recognition* 11:19-28.
- Selwyn, E. W. H. 1939. Experiments on the nature of graininess. *Photogr. J.* 79:513-524.
- Selwyn, E. W. H. 1959. The 12th Renwick Memorial lecture: Scientists' ways of thinking about definition. *J. Photogr. Sci.* 7:138-147.
- Stansfield, S. A. 1986. ANGY: A rule-based expert system for automatic segmentation of coronary vessels from digital subtracted angiograms. *IEEE Trans. Pattern Anal. Machine Intell.* PAMI-8:188-199.
- Suresh, B. R., and B. A. Shenoi. 1981. New results in two-dimensional Kalman filtering with application to image restoration. *IEEE Trans. Circuits Syst.* CAS-28:307-319.
- Tekalp, A. M., H. Kaufman, and J. W. Woods. 1986. Identification of image and blur parameters for the restoration of noncausal blurs. *IEEE Trans. Acoust Speech Signal Processing* ASSP-34:963-972.
- Tou, J. T., and R. C. Gonzalez. 1974. *Pattern Recognition Principles*. Addison-Wesley, Reading, Mass., 1974.
- Woods, J. W., and C. Radewan. 1977. Kalman filtering in two dimensions. *IEEE Trans. Inf. Theory* IT-23:473-482.
- Zheng, Y., and J. P. Basart. 1988. NDE x-ray image modeling and adaptive filtering considering correlated noise. In D. O. Thompson, and D. E. Chimenti, ed. *Review of Progress in Quantitative NDE*. Vol. 7 (in press). Plenum Press, New York.

Zheng, Y., and J. P. Basart. 1987. Automatic image segmentation, modeling, and restoration with a rule-based expert system. Pages 421-425 in D. M. Etter, ed. Twentieth Asilomar Conference on Signal, System & Computers. IEEE Inc., New York.

PART III. APPLICATION OF ADAPTIVE REGIONAL KALMAN FILTERING TO X-RAY
IMAGES IN NDE

ABSTRACT

X-ray images of flaws in thick dense materials can be improved by image enhancement techniques. These images are usually noisy and have low contrast. We have developed a procedure for enhancing the contrast of a flaw and smoothing the image while preserving the edges of the flaw. The steps involved in this procedure are: 1) image segmentation, 2) autoregressive (AR) model identification, and 3) adaptive regional Kalman filtering. An image is segmented by a modified K-mean clusterseeking algorithm. This algorithm uses information from local means and from local masking functions representing spatial activity of an image. Each of the segments is divided into many small windows. After this, each window is modeled by an AR process. By keeping each window small, e.g., 7 X 7 pixels, the model more clearly represents the data. The model of each window is transformed to a state-space form and put into a Kalman filter. Each window is smoothed four times by passing the Kalman filter through a central pixel of interest from four different orthogonal directions. The four passes giving four estimates of the pixel intensity, are optimally weighted and combined to give the final pixel estimate. This enhancement procedure includes several advancements in image processing by emphasizing local information, incorporating noncausality, preserving edges, and adaptively identifying and estimating regional models. Many image enhancement techniques will smooth out the noise, but they also blur the edges and the details of the image.

INTRODUCTION

One of the time-consuming procedures in inspecting parts by x-ray film is the identification of a flaw. Low contrast films of dense objects especially cause problems. A radiologist must have considerable experience in identification in order to keep the examination time relatively small, but also keep the reliability high. Our objective in this project is to develop a computer procedure that will sufficiently enhance flaws in an image in a manner that will reduce the time it takes a human to locate and identify a flaw.

Factors limiting the quality of an X-ray image are image unsharpness, quantum fluctuation, film grain and film contrast [Halmshaw, 1973; 1982]. The unsharpness caused by scattered radiation reduces the image contrast. The quantum fluctuation caused by random emission and absorption of X-ray quanta smears or masks the contrast. The film grain and contrast limit the recorded information capacity. A coarse-grained image conveys less detail than one of fine grain. In this paper, we discuss a method for enhancing the image by reducing the fluctuation due to disturbances, such as quantum fluctuation and granularity, etc. The main tool used is the Kalman filter. The basic idea is to estimate a pixel optimally in an image using a given pixel and its near neighbors. The Kalman filter can include a model of the process that generated the desired information, a model of the noise added to this process, a model of the measurement system, and a model of the noise within the measurement system. In addition, there can be

multiple models representing multiple processes at any one, or all, of these stages. Another advantage of the Kalman filter is that it can distinguish between stochastic processes that have strongly overlapping spectra. Ordinary spectral filters are of limited benefit under such conditions. When processing noisy images one often finds that the noise, system, and signal processes overlap in frequency.

Three steps are involved when implementing our method of filtering. They are 1) image segmentation, 2) image modeling, and 3) Kalman filtering. Each of these procedures will be explained. The results of filtering a low contrast flaw in an x-ray image will be discussed at the end of the article.

SEGMENTATION

Segmentation is an image classification procedure. Autoregressive modeling, which we incorporate in our method, requires stationarity. Generally, the stationarity assumption is not true for the processes in an image over the whole image and this violation will cause blurred edges and reduced contrast in a filtered image. Therefore segmentation is necessary to find regions in which the statistics, mean and variance, are stationary.

An image is segmented by partitioning it with respect to local mean and local spatial activity of the image [Schachter et al., 1979; Rajala and de Figueiredo, 1981]. Spatial activity is defined as the rate of change of spatial luminance from one pixel to another. It is related to the concept of variance. The formula used to calculate the spatial activity is called the masking function. Regions of stationary mean and stationary variance can be found by segmenting an image by local means and by the masking function, respectively. With these two segmentations in hand, they can be combined to produce new segments that are wide-sense stationary.

Local means are found by a window of running average. A $(2n+1) \times (2n+1)$ window is selected in one corner of the image. All pixels within the window are averaged. This average is assigned to the center pixel. The window is then moved and the process is repeated. The mathematical expression for the local mean is

$$m_n(i, j) = \frac{1}{(2n+1)^2} \sum_{p=i-n}^{i+n} \sum_{q=j-n}^{j+n} z(p, q) \quad (3.1)$$

where $z(p, q)$ is the image intensity at pixel p, q . After calculation of the mean for all windows, a file of the local means is set aside for later use.

The next step is to determine the masking function for the image. The masking function is defined by

$$M_r(i, j) = \sum_{p=i-r}^{i+r} \sum_{q=j-r}^{j+r} e^{-|| (x, y) - (p, q) ||} \left[(1/4) \sum_{p=0}^3 D_{pqn} \right] \quad (3.2)$$

where $|| (x, y) - (p, q) ||$ is the Euclidean distance between points (x, y) and (p, q) , (x, y) is center pixel of a window, (p, q) is any other point in the window, and D is the difference in intensity between a pixel adjacent to (p, q) and the pixel at (p, q) . The difference, D , is summed over all pixels adjacent to (p, q) . The average of these differences is weighted exponentially by the distance from (p, q) to (x, y) . After the masking function is calculated for all the windows, it is recorded in a file.

The next step is to use the local means and masking function to segment the image. A cluster seeking procedure, somewhat similar to the K-means cluster seeking algorithm [Tou and Gonzalez, 1974], is used to cluster local means and masking functions. It differs from the standard K-means cluster seeking algorithm in that the thresholds of

the distance between the cluster center are given for simplicity. Each local mean and masking function is assigned to a certain cluster. All combinations of local mean clusters and masking clusters form wide-sense stationary regions which we desire.

MODELING

After completing the segmentation, the process in each segmented region is represented by a p-order AR process:

$$s(k) = \sum_{n=1}^p \phi_n s(k-n) + w(k) \quad (3.3)$$

$$z(k) = H s(k) + v(k) \quad (3.4)$$

where $z(k)$ is the measurement of intensity at a pixel, $v(k)$ is an additive measurement white-noise sequence, $s(k)$ is a "true image" process, $w(k)$ is a residual sequence, H is a $(1 \times m)$ measurement vector and $s(k)$ is an $(m \times 1)$ vector of $s(k)$. $v(k)$ and $w(k)$ are independent and uncorrelated with $E[v(k)] = 0$, $E[w(k)] = 0$, $E[v(k)w(h)] = 0$, $E[v(k)v(k-h)] = R \cdot \delta(h)$ and $E[w(k)w(k)] = Q \cdot \delta(h)$. ϕ 's are coefficients to be estimated. There are a number of ways to estimate ϕ 's such as maximum likelihood or least squares approaches [Box and Jenkins, 1976]. "Marquardt's compromise" [Marquardt, 1963; Pankratz, 1983] and Yule-Walker equation [Box and Jenkins, 1976] methods are often used in practice. We estimate the ϕ 's by solving the Yule-Walker equation

$$r \phi = r_1 \quad (3.5)$$

where

$$\phi = [\phi_1 \quad \phi_2 \quad \phi_3 \quad . \quad . \quad . \quad \phi_p]^T$$

$$\mathbf{r}_1 = [r_1 \ r_2 \ r_3 \ \dots \ r_p]^T$$

$$\mathbf{r} = \begin{bmatrix} r_0 & r_1 & \dots & r_{p-1} \\ r_1 & r_0 & \dots & \cdot \\ \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot \\ r_{p-1} & \cdot & \dots & r_0 \end{bmatrix}$$

The r 's are autocorrelation coefficients of $s(k)$. Given the measured $z(k)$'s and the variance R of $v(k)$, the r 's can be found by taking the expectation of (Rajala and de Figueiredo, 1981). The semi-positive definite property of the r 's must be considered when the r 's are calculated [Fuller, 1976].

After a state-space form of (3.3) is obtained [Abraham and Ledolter, 1983], we are ready to apply the Kalman filter.

KALMAN FILTERING

The Kalman filter is an optimal linear filter that can separate two or more stochastic process. The Kalman filter theory and applications can be found in many sources [Kalman, 1960; Kalman and Bucy, 1961; Brown, 1983].

Since all quantities required for Kalman filtering have now been found, the optimal estimates of pixels are obtained by the following recursive procedure:

1. Enter the recursive loop with the initial values of the a prior estimated (nx1) vector $s(k|k-1)$ and its error covariance matrix $P(k|k-1)$
2. Compute the Kalman gain

$$K(k|k) = P(k|k-1) H^T (H P(k|k-1) H^T + R)^{-1} \quad (3.6)$$

3. Estimate a pixel

$$s(k|k) = s(k|k-1) + K(k|k) (z(k) - H s(k|k-1)) \quad (3.7)$$

4. Compute the error covariance matrix

$$P(k|k) = (I - K(k|k) H) P(k|k-1) \quad (3.8)$$

5. Predict

$$s(k+1|k) = \Phi s(k|k) \quad (3.9)$$

$$\mathbf{P}(k+1|k) = \Phi \mathbf{P}(k|k) \Phi^T + Q$$

Φ is a transition $p \times p$ matrix from equation (3.3) [Abraham and Ledolter, 1983]. The process is repeated for the next pixel $z(k+1)$ from step 2 until all pixels are processed. One should be careful that the Kalman equations are simplified due to the scalar modeling.

RESULTS

By applying the above procedure to low-contrast X-ray images, we have produced enhanced images. One example of a processed image is shown here. It is an 88x88 pixel subimage of an X-ray image of a casting. There is a flaw located near the center area of the image. The flaw is not obvious in the original image which is very dense and has low contrast. The contour map of the original image is shown in Figure 3.1(a). The variance of the disturbance fluctuation measured from a flat area in the original image is 1.6. The result from filtering is shown in the contour map in Figure 3.1(b). Since the dynamic ranges of the images are too small (about 20 to 30), histogram equalization with an exponential transformation function was applied to both the original and filtered images. Ruled surface plots of the original and the transformation results are shown in Figure 3.2(a) and Figure 3.2(b), respectively. The dynamic range in Figure 3.2 has increased to 128. In the original image, the flaw region is broken into many spikes which make flaw detection difficult. The filtered image shows a bigger concentration of intensity within a region that can be defined by a single boundary. Compared with the original image, the filtered one has a lower and smoother background. Thus the flaw in the filtered image is easily detected now.

The experiment was done on an ISU VAX 11/780 computer. The CPU time for running the Kalman filter part was about 8 minutes (88x88 pixels) using a moving window. The modeling and filtering were applied

to a 7x7 moving window and the 3x3 pixels in the center of the window were saved each time.

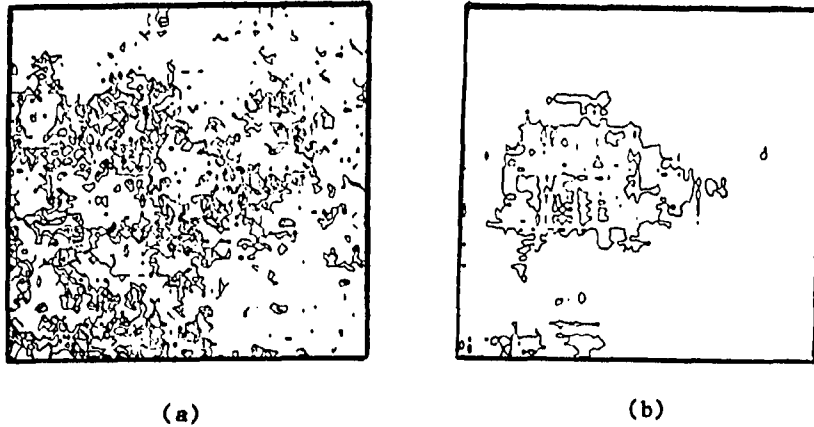


Figure 3.1. Contour plots of intensity before (a) and after (b) filtering. Contour levels are the same in both plots. The flaw in the filtered map (b) clearly stands out above the background.

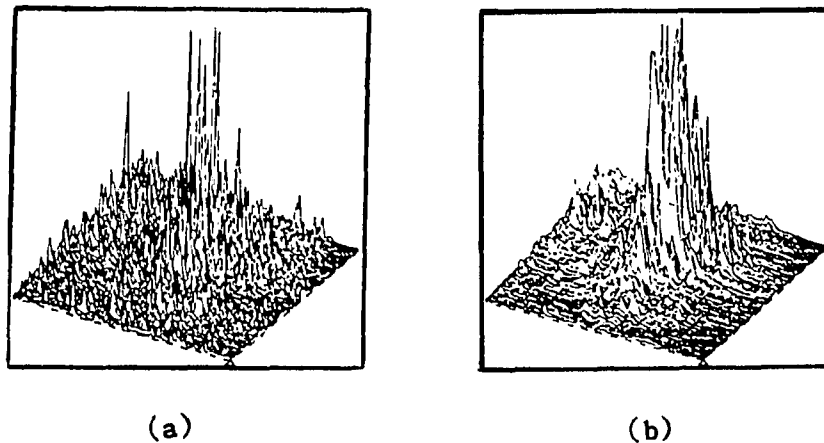


Figure 3.2. Ruled surface plots before (a) and after (b) filtering. In the filtered map (b), the background noise is lowered and smoothed, and the power in the flaw is more centralized and less "spikey" than in the original map.

ACKNOWLEDGMENT

This work was supported by NSF grant No. AST-8217135 and the NSF University/Industry Center for NDE at Iowa State University.

REFERENCES

- Abraham, B. and J. Ledolter. 1983. Statistical methods for forecasting. John Wiley and Sons, New York.
- Box, G. E. P., and G. M. Jenkins. 1976. Time series analysis: forecasting and control. Holden-Day, San Francisco, CA.
- Brown, R. G. 1983. Introduction to random signal analysis and Kalman filtering. John Wiley & Sons, New York.
- Fuller, W. A. 1976. Introduction to Statistical Time Series. John Wiley and Sons, New York.
- Halmshaw, R. 1982. Industrial radiology: theory and practice. Applied Science Publishers Ltd, Englewood, New Jersey.
- Halmshaw, R. 1973. Information theory applied to industrial radiography. Pages 31-64 in R. S. Sharpe, ed. Research techniques in nondestructive testing. Vol 2. Academic Press, New York.
- Kalman, R. E. 1960. A new approach to linear filter and prediction problems. Trans. ASME 82:35-45.
- Kalman, R. E., and R. S. Bucy. 1961. New results in linear filtering and prediction theory. Trans. ASME 83:95-108.
- Marquardt, D. W. An algorithm for least squares estimation of non-linear parameters. 1963. J. Soc. Ind. Appl. Math. 11:431-441.
- Pankratz, A. 1983. Forecasting with univariate Box-Jenkins models. John Wiley, New York, NY.
- Rajala, S. A., and R. I. P. de Figueiredo. 1981. Adaptive nonlinear image restoration by a modified Kalman filtering approach. IEEE Trans. Acoust., Speech, Signal Processing ASSP-29:1033-1042.
- Schachter, B. J., L. S. Davis, and A. Rosenfeld. 1979. Some experiments in image segmentation by clustering of local feature values. Pattern Recognition 11:19-28.
- Tou, J. T., and R. C. Gonzalez. 1974. Pattern Recognition principles. Addison-Wesley, Reading, MA.

PART IV. AUTOMATIC IMAGE SEGMENTATION, MODELING AND RESTORATION WITH A
RULE-BASED EXPERT SYSTEM

ABSTRACT

A noisy image can be improved by an adaptive filtering procedure. A procedure we have developed includes four major steps: 1) low-level image segmentation, 2) noise measurement, 3) ARMA image model identification and estimation, and 4) adaptive Kalman filtering. This procedure smooths out noise but preserves edges and the details of image well. The procedure is controlled by an expert system with no human interaction. Little a priori knowledge about the "true image" is required. General knowledge (as rules) about the whole procedure and a set of control rules are used in the rule-based expert system.

INTRODUCTION

Image segmentation, modeling, and restoration are subjects of great interest in image processing. They find wide application in a variety of areas such as medical thermography, nondestructive evaluation, radio astronomy, synthetic radio imagery, satellite imagery, and military target screening. They are often used for picture analysis, feature extraction, target identification, and noisy image enhancement. Much research progress has been accomplished and many algorithms have been developed for these subjects. However, there often can be a gap between the invention of an algorithm and the development of the algorithm for useful practice since the amount of knowledge required to use new methods can be extensive. In our research we are developing image processing algorithms for enhancing X-ray images in the area of nondestructive evaluation (NDE) and for improving radio astronomical images created by aperture synthesis. In this paper, we report our initial results of implementing an automatic realization of our image processing procedure by a rule-based expert system.

Our procedure, which includes image segmentation, noise measurement, modeling, and Kalman filtering, is designed to restore a noisy image (degradation by Gaussian white noise is assumed at this stage of our work). In the last decade, a Kalman-filter approach to image restoration has been given such great attention that much progress has been made [Woods and Radewan, 1977; Woods 1981; Suresh and

Shenoi, 1981; Rajala and de Figueiredo, 1981; Dikshit, 1982; Biemond et al., 1983]. For most of these approaches, stationarity of an image, or of a strip, or of a window has been assumed for the convenience of processing the image. In addition, models of original images are assumed in some approaches. The assumptions of stationarity and of an available model are often not true. Stationarity may not be valid due to the abrupt variation of an image. Complete a priori knowledge about an original image may not be available in practice. In our approach, little a priori knowledge of the original image is required. The assumptions are 1) the noisy image is degraded by Gaussian white noise with an unknown variance, and 2) there is a moderately-sized flat area or an area whose trend can be removed in the image. Low-level image segmentation is used to partition an image into approximately wide-sense stationary areas. Then modeling and Kalman filtering are applied to each irregularly segmented area. Since the filtering procedure is realized in each segmented area, edges and details of an image can be accurately preserved. Because irregular areas are considered, a region-seeking and scanning algorithm is necessary. Monitoring and controlling of the procedure are achieved by a region growing algorithm, an area scanning algorithm, and a tree-structure seeking algorithm. These three main control rules and the low-level properties of an image are used as knowledge rules in the rule-based expert system.

ALGORITHMS

The three steps discussed in the following material are 1) image segmentation, 2) noise estimation and image modeling, and 3) adaptive Kalman filtering.

Image Segmentation

Image segmentation is a process of classification. Image classification consists of low-level and high-level evaluations and has been widely used in feature extraction. Low-level segmentation consists of studying an image and understanding the processes involved. In this stage, an image is partitioned into uniform regions according to spatial activities of the image [Rajala and de Figueiredo, 1981; Schachter et al., 1979]. Whereas, at the high-level stage one interprets the natural contents of an image. In this paper we consider only low-level segmentation. Intensity, contrast, edges, sizes, adjacency, and the shape of an image, provide information for segmenting an image.

An image is primarily segmented by the histogram of the local mean. The local mean is found by a $(2n+1) \times (2n+1)$ window of running average. The local mean at pixel (i,j) is

$$m_n(i,j) = \frac{1}{(2n+1)^2} \sum_{p=i-n}^{i+n} \sum_{q=j-n}^{j+n} z(p,q) \quad (4.1)$$

where $z(p,q)$ is the image intensity at pixel (p,q) . The thresholds separating modes in a histogram of local means are then calculated by a local minimum-value seeking algorithm. This algorithm finds local minimum values by searching for slope changes in a smoothed histogram. Local minimum values caused by local fluctuations are trapped and ignored.

The spatially activities and edges are described by the masking function which is defined as

$$M_r(i,j) = \sum_{p=i-r}^{i+r} \sum_{q=j-r}^{j+r} [e^{-|| (i,j)-(p,q) ||}] \cdot [(1/4) \sum_{n=0}^3 D_{pqn}] \quad (4.2)$$

where $|| (i,j)-(p,q) ||$ is the Euclidean distance between point (i,j) and (p,q) , and D is the slope at the pixel (p,q) . To reduce the effects of noise, two or three pixels adjacent to (p,q) at each side are used to calculate the slope for one of four directions. The standard masking function uses one adjacent pixel at each side to find D , and is very sensitive to noise.

A set of decision rules is then made to segment an image. We introduce this in the next section.

Noise Measurement and Image Modeling

If we have satisfactorily segmented an image, then the process in each segmented region is approximately wide-sense stationary. An irregular area seeking and scanning (as a control rule) algorithm is

used to collect data into an one-dimensional array which can be represented as

$$s(k) = \sum_{n=1}^p \phi_n s(k-n) + \sum_{n=1}^q \theta_n w(k-n) + w(k) \quad (4.3)$$

where $z(k)$ is the measured intensity at a pixel, $v(k)$ is an additive measurement white-noise sequence, $s(k)$ is an original image process represented as a ARMA(p,q), and $w(k)$ is a residual sequence. The sequences $v(k)$ and $w(k)$ are independent and uncorrelated with $E[v(k)] = 0$, $E[w(k)] = 0$, $E[v(k)w(k)] = 0$, $E[v(k)v(k-h)] = R\delta(h)$, and $E[w(k)w(k-h)] = Q\delta(h)$. Since only the noisy image is available, we have to find the noise level from the available image. Theoretically, if there is a moderately sized flat region in which pixels of an original image are close to a constant, the variance of the noise can be found by calculating the variance of the noisy image in this region. We developed an algorithm to search for this area. Since the local mean is not sensitive to noise, the algorithm seeks this region from the segmented local mean file. If all segmented areas of the local mean do not meet the requirements, a failure signal is given. The requirement for success is that the variance of an segmented region of local mean has to be less than a certain value.

The next step is to find the coefficients of equation (4.3) given only the measurement sequence $z(k)$. There are two methods for doing this. One of them finds an AR(p) process of $z(k)$ by solving the Yule-Walker equation [Box and Jenkins, 1976], and substitutes this model

into equation (4.4). Then an ARMA(p,q) form of $s(k)$ can be found. The coefficients of the AR part of $s(k)$ are the same as those of the AR(p) model of $z(k)$, and the coefficients of the MA part of $s(k)$ can be found by solving a set of non-linear equations which could be complicated except for the case of $p=1$. However, an ARMA(1,1) model can represent many processes. Another method for finding the coefficients of $s(k)$ involves subtracting the noise variance from the covariance of $z(k)$, and then finding an ARMA($p,0$) model of $s(k)$ by solving the Yule-Walker equation. In the second method, one must adhere to the semi-positive definite of the autocorrelation coefficients of $s(k)$ [Fuller, 1976]. Hence a factor whose value is less one is multiplied by the noise variance before the subtraction. In both methods, an equation for calculating the Q value is needed. The ARMA model can be written in a state-space form and implemented with the Kalman filter [Abraham and Ledolter, 1983].

Adaptive Kalman Filtering

A bank of Kalman filters is used to separate the original image from the noise. The Kalman filter is an optimal linear filter which distinguishes between two or more additive stochastic processes whose spectra overlap. The Kalman filter is applied to each segmented region with a different state transition matrix obtained in the modeling stage. Therefore, the Kalman filter parameters are adapted to the local features of an image. Kalman filter theory has been introduced

in many textbooks and papers [Kalman, 1960; Kalman and Bucy 1961; Brown, 1983] so we do not list the recursive equations here.

RULE-BASED EXPERT SYSTEM

"An expert system is a rule-based AI application program for doing a task which requires expertise" [Charniak and McDermott, 1985]. In recent years, rule-based expert systems have been applied to image segmentation [Nazif and Levin, 1984; Stansfield, 1986]. The general structure of an expert system can be described by three major components: 1) a short-term memory (STM), 2) a long-term memory (LTM), and 3) an interpreter. The STM stores input images, local means, masking functions, histogram of local means, segmentation marks of segmented regions, data tree structure and leaves, data stacks and pointers, and intermediate results of the processed image. The LTM stores the knowledge about low-level segmentation, modeling, and the procedural control strategies. Knowledge is represented by condition-action rules of the form:

```

FROM      <phenomenon>
IF        <condition 1 .AND. condition 2 .AND. ... .AND.
          condition N .OR. condition N+1 .OR. ... .OR.
          condition N+M>,
THEN      <actions>

```

Some of the rules we use are:

A. Rules for Merging

RULE (201):

FROM <Dynamic range of a region is too small>
 IF <The region size is very low .AND. adjacency with
 another region is high .AND. the difference in
 local mean with another region is not high .AND.
 the difference in masking function with another
 region is not high>
 THEN <Merge the two regions>

RULE (202):

FROM <The region size is small>
 IF <The region size is higher than a certain value
 .AND. adjacency with another region is high .AND.
 the difference in local mean with another region is
 not high .AND. the difference in masking function
 with another region is not high>
 THEN <Merge the two regions>

RULE (203):

FROM <The region size is too small>
 IF <Adjacency with another region is high .AND. the
 difference in local mean with another region is not
 high .AND. the difference in masking function with
 another regions is not high>
 THEN <MERGE THE TWO REGIONS>

RULE (204):

FROM <The region size is less than 3>

IF <There is only one adjacent region>
 THEN <Merge the two regions>

RULE (205):

FROM <The region size is less than 3>
 IF <The difference between the mean value and the
 center of adjacent region I is a minimum among
 other adjacent regions>
 THEN <Merge this region with region I>

B. Rules to Find segmentation Threshold in a Histogram

RULE (301):

FROM <There is a double mode in the histogram>
 IF <The difference between centers of the modes is
 higher than a constant + $2 \times \text{rms}$ of the noise>
 THEN <Choose a value which has an equal distance to the
 two centers as a segmentation threshold>

RULE (302):

FROM <There is a local minimum value>
 IF <The local peaks in both sides are very high .AND.
 the distance between the two peaks is high>
 THEN <Choose this local minimum value as a segmentation
 threshold>

RULE (303):

FROM <There is a slope change in the smoothed histogram>

IF <The ratio between a location where the slope
 changes and an adjacent peak is large .AND. the
 slope changes from minus to plus>
 THEN <The location is a local minimum>

RULE (304):

FROM <There is a slope change in the smoothed histogram>
 IF <The ratio between a peak and an adjacent minimum is
 large>
 THEN <Define the minimum as a segment boundary>

C. Rules for Segmentation

RULE (401):

FROM <The region size is large>
 IF <There is a double mode .AND. the dynamic range is
 high>
 THEN <Segment this region by a threshold>

RULE (402):

FROM <The dynamic range of a region is high>
 IF <There is a double mode>
 THEN <Segment this region according to the histogram>

RULE (403):

FROM <The variance of a region is high>
 IF <The region histogram is bimodal>

THEN <Segment this region>

RULE (404):

FROM <Region histogram is bimodal>

IF <The region size is not small>

THEN <Segment this region>

RULE (405):

FROM <Region size is medium>

IF <The region histogram bimodality is very high>

THEN <Segment this region according to the histogram>

D. Rule for Noise Measurement:

RULE (501):

FROM <Region size is moderate>

IF <The variance of local means in this region is not
 high>

THEN <The value of ((variance of the region of the noisy
 image) - (variance of the region of the local
 means)) is an approximation of the additive white
 noise variance>

Other rules for region growing and area seeking are not listed here for the sake of brevity.

EXPERIMENTAL RESULTS

The procedure has been tested with a simulation. Figure 4.1 is a digitized version of a photograph of the Very Large Array (VLA) radio telescope operated by the National Radio Astronomy Observatory in Socorro, New Mexico. The size of the digitized image is 480×512 . The intensity range is 0 to 255. The noisy image created from Figure 4.1 is shown in Figure 4.2. The variance of the additive noise is 100.0. Figure 4.3 is a picture of the local means made from Figure 2 which is similar to the result of a mean filter. The edges are blurred in the local-mean picture. Figure 4.4. is a result of applying a gradient method to detect the edges of the noisy image in Figure 2. Figure 4.4 shows that the gradient method is very sensitive to noise. Figure 4.5 is a result of applying the modified masking function to the noisy image which retains the definition of the edges. Figure 4.6 is one of the segments and Figure 4.7 shows an edge-detected version of Figure 4.6. It is interesting to note that the features of the antennas are clearly shown. Figure 4.8 is a Kalman filtered image from Figure 4.2. The figure shows that the Kalman filter smooths out noise and preserves the edges well. The noise variance is reduced by about 4 dB. It took about 20 minutes of CPU time on an NAS-9 computer to run this procedure for a 480×512 image. We are developing a modified version that may reduce the computation time to about 5 to 7 minutes of a NAS-9 CPU time.

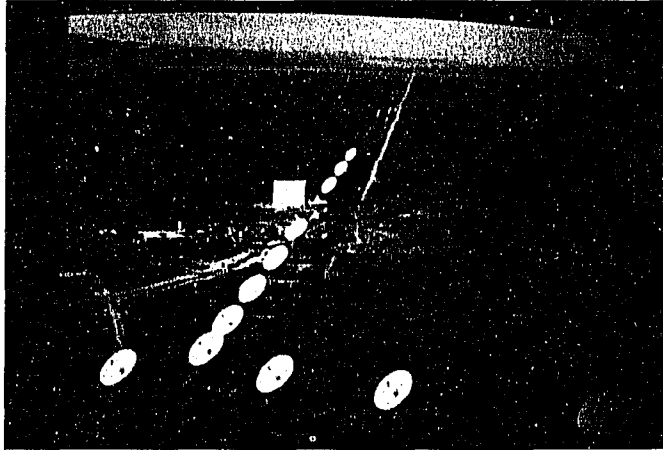


Figure 4.1. Original digitized image

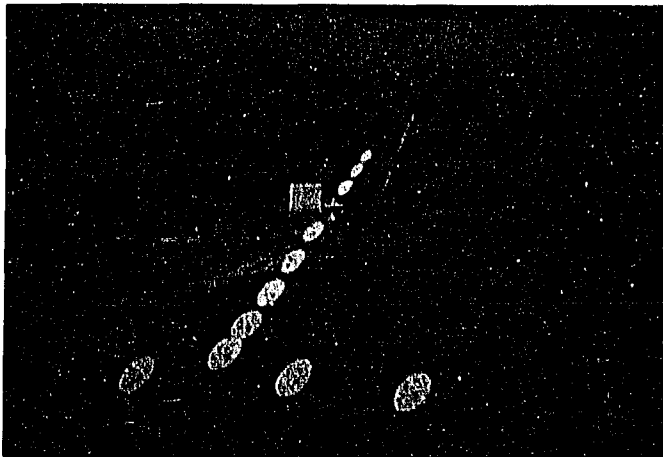


Figure 4.2. Image with noise

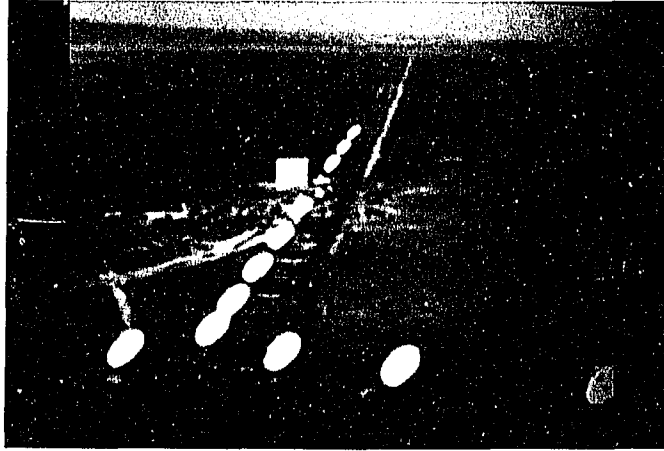


Figure 4.3. Image of local means of Figure 4.2

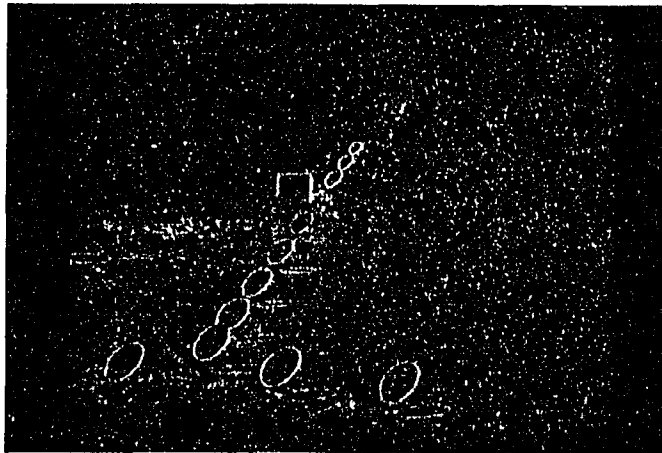


Figure 4.4. Image of gradients of Figure 4.2

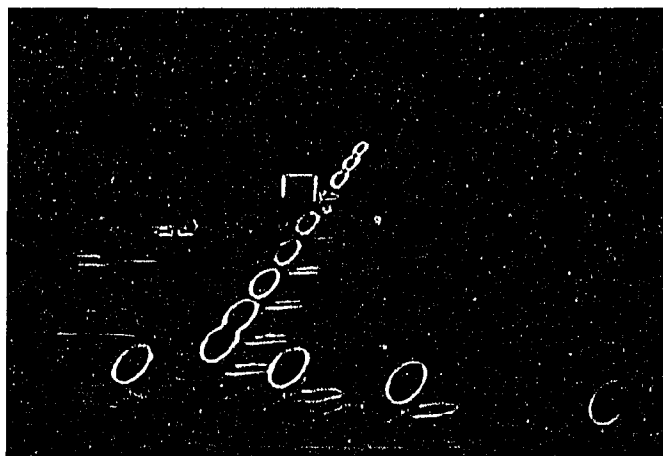


Figure 4.5. Image of modified masking function of Figure 4.2

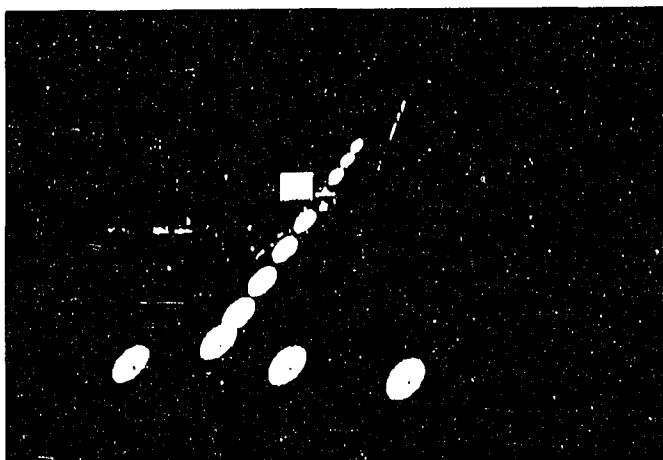


Figure 4.6. Image of one segment

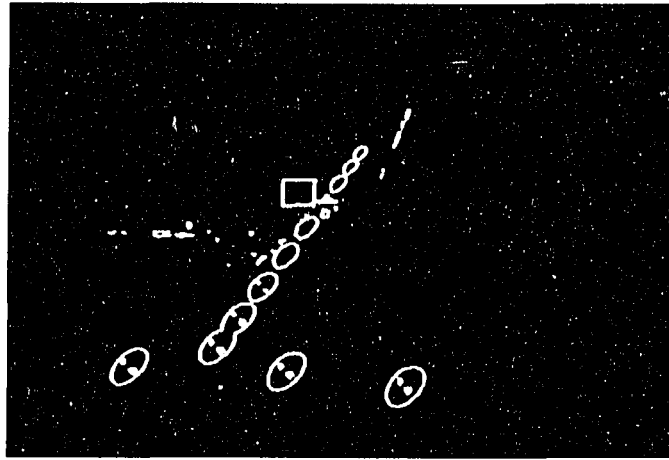


Figure 4.7. Edge detected version of Figure 4.6

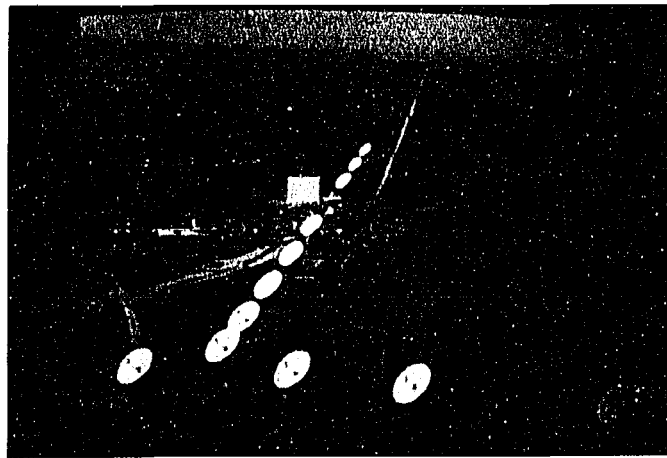


Figure 4.8. Final Kalman filtered image

ACKNOWLEDGMENTS

This work was supported by the Center for Nondestructive Evaluation at Iowa State University and by National Science Foundation grant no. AST-8217135.

REFERENCES

- Biernond, J., J. Rieske, and J. J. Gerbrands. 1983. A fast Kalman filter for images degraded by both blur and noise. IEEE Trans. Acoust., Speech, Signal Processing ASSP-31:1248-1256.
- Abraham, B. and J. Ledolter. 1983. Statistical methods for forecasting. John Wiley and Sons, New York.
- Box, G. E. P., and G. M. Jenkins. 1976. Time series analysis: forecasting and control. 2nd ed. Holden-Day, San Francisco, CA.
- Brown, R. G. 1983. Introduction to random signal analysis and Kalman filtering. John Wiley & Sons, New York.
- Charniak, E. and D. McDermott. 1985. Introduction to artificial intelligence. Addison-Wesley, Reading, MA.
- Dikshit, S. S. 1982. A recursive Kalman window approach to image restoration. IEEE Trans. Acoust., Speech, Signal Processing ASSP-30:125-140.
- Fuller, W. A. 1976. Introduction to Statistical Time Series. John Wiley and Sons, New York.
- Kalman, R. E. 1960. A new approach to linear filter and prediction problems. Trans. ASME 82:35-45.
- Kalman, R. E., and R. S. Bucy. 1961. New results in linear filtering and prediction theory. Trans. ASME 83:95-108.
- Nazif, A. M. and M. D. Levine. 1984. Low level image segmentation: An expert system. IEEE Trans. Pattern Anal. Machine Intell. PAMI-6:555-577.
- Rajala, S. A., and de Figueiredo, R. I. P. 1981. Adaptive nonlinear image restoration by a modified Kalman filtering approach. IEEE Trans. Acoust., Speech, Signal Processing ASSP-29:1033-1042.
- Schachter, B. J., L. S. Davis, and A. Rosenfeld. 1979. Some experiments in image segmentation by clustering of local feature values. Pattern Recognition 11:19-28.
- Stansfield, S. A. 1986. ANGY: A rule-based expert system for automatic segmentation of coronary vessels from digital subtracted angiograms. IEEE Trans. Pattern Anal. Machine Intell. PAMI-8:188-199.

- Suresh, B. R., and B. A. Shenoi. 1981. New results in two-dimensional Kalman filtering with application to image restoration. IEEE Trans. Circuits Syst. CAS-28:307-319.
- Woods, J. W. 1981. Two-dimensional kalman filtering. Pages 155-205 in T. S. Huang ed. Topics in Applied Physics. Vol. 42. Springer-Verlag, Berlin.
- Woods, J. W., and C. H. Radewan. 1977. Kalman filtering in two dimensions. IEEE Trans. Inform. Theory IT-23:473-482.

PART V. NDE X-RAY IMAGE MODELING AND ADAPTIVE FILTERING CONSIDERING
CORRELATED NOISE

ABSTRACT

A technique for modeling and adaptive filtering with consideration of colored noise was developed to increase the contrast of NDE X-ray images. It can be statistically shown that noise limiting the thickness sensitivity of an NDE X-ray image is spatially correlated. This knowledge supports an AR representation of noise. The noise model and statistics of a noisy image are used to find an AR signal model. Signal and noise models can be transformed to state-space forms which represent an image. This image representation is closer to a practical situation than that for a white noise assumption. Incorporated with the image model, a Kalman filter is used to enhance the image by optimally separating the signal and noise processes. In this case, colored noise is modeled as well as the signal in the Kalman filter equations. Since we used a fast computation method and took advantage of a recursive property, it was possible to incorporate a colored noise model in the Kalman filter without significantly increasing the computation time. It takes about one minute (not including I/O time) of VAX 11/780 CPU time to filter a 200x200 image with a second order Kalman filter, which can be comparable to or even less than the computation time of many simple smoothing algorithms.

INTRODUCTION

Four main factors limiting the quality of an industrial X-ray radiographic image are unsharpness, quantum fluctuation, film-grain noise and film contrast [Halmshaw, 1973; 1982]. Unsharpness blurs details of an image, film contrast limits the dynamic range of an image, and quantum fluctuation and film-grain noise smear the fine structure of a specimen and reduce the contrast of an image. The effects of these factors degrading the quality of an X-ray image are well discussed in several books [Halmshaw, 1982; Herz, 1969; Dainty and Shaw, 1974].

Techniques for reducing the quantum fluctuation and the film-grain noise have been studied by Kuan et al. who developed an adaptive smoothing filter to reduce the signal-dependent film-grain noise [Kuan et al., 1985], by Naderi and Sawchuk who developed an adaptive estimation filter, based on a discrete Wiener filter, to restore an image degraded by film-grain noise [Naderi and Sawchuk, 1978], and by Basart et al. who used an adaptive regional Kalman filter to remove the quantum fluctuation in industrial-radiographic X-ray images [Basart et al., 1987].

In this work, an adaptive Kalman filtering technique is developed to remove noise fluctuation caused by film-grain noise, quantum fluctuation, and film dirt of industrial-radiographic X-ray images. With this technique, knowledge of a radiographic X-ray image forming mechanism is employed to obtain Autoregressive (AR) models of an image.

The AR representation of an image is written in a state-space form and then incorporated into a Kalman filter that enhances an image by optimally separating a noise-free image process from a noise process. Since the image model of this technique closely represents a real image, a better estimation of an image can be obtained than that of many other methods. It can be shown that the adaptive filtering schemes given by [Kuan et al., 1985] and [Basart et al., 1987] are special cases of this technique. In addition, since the Kalman filter recursively processes data, the execution speed of this method is fast. The computation time is comparable to that of many simple smoothing algorithms.

X-RAY RADIOGRAPHIC IMAGE MODELING

To meet the objective of removing noise fluctuation, the image forming mechanism must be studied and modeled. An X-ray image forming mechanism is a very complicated optical and chemical process. Various efforts have been made to obtain proper mathematical descriptions of this process [Halmshaw, 1973; 1982; Dainty and Shaw, 1974; Gray, 1988; Higgins and Stultz, 1959]. A highly accurate and complete description may cause difficulty when one tries to derive a practical restoration scheme for an image. Therefore, a simple and reasonably accurate model is desired.

Let a class of X-ray radiographic images be characterized by a discrete real-valued finite random field $X = \{X_{ij}; 1 \leq i \leq N_1, 1 \leq j \leq N_2\}$. A realization of this field is $x = \{x_{ij}; 1 \leq i \leq N_1, 1 \leq j \leq N_2\}$ where x_{ij} is an intensity value at pixel (i, j) of a specific image. Assuming that the film is processed in the linear region of the D-logE curve and ignoring the blurring effect, a realization of the observation field can be represented by [Kuan et al., 1985; Naderi and Sawchuk, 1978; Huang, 1966]

$$y(i, j) = x(i, j) + v_1(i, j) \quad (5.1)$$

where

$$v_1(i, j) = k f(x(i, j)) v_2(i, j). \quad (5.2)$$

the term v_1 is signal-dependent film-grain noise with a Gaussian distribution $N(0, k^2(x(i,j))^{1/3})$. The function $f(x(i,j))$ indicates that $v_1(i,j)$ is signal-dependent noise. $v_2(i,j)$ is $iid(0,1)$, and k is a factor related to a ratio of a mean film-grain size to the size of a scanning aperture of a digitization system. Film-grain noise refers to a natural statistical fluctuation of randomly distributed silver particles. A measurement of this fluctuation is also given by Selwyn's law [Selwyn, 1939; 1959]. In general, film-grain noise is spatially correlated [Dainty and Shaw, 1974], but the correlation distance is often very short compared to the scanning aperture size of a digitization system. Readers should notice that there is a confusion about this equation in the literature given by Huang and Kuan et al. [Kuan et al., 1985; Huang, 1966]. They represented $v_1(i,j)$ by $k(x(i,j))^{1/3}v_2(i,j)$. If we treat $x(i,j)$ as a random variable, the variance of $v_1(i,j)$ obtained from their representation will be different than that mentioned above, $k^2(x(i,j))^{1/3}$.

equation (5.1) is very important because it reveals the relation between the optical density and the film-grain noise. However, Naderi and Sawchuk showed that equation (5.1) is not accurate enough to simulate the noise fluctuation of a real X-ray radiographic image [Naderi and Sawchuk, 1978]. The noise fluctuation generated by equation (5.1) has a fine granular salt and pepper appearance, whereas the noise is smooth in a real X-ray image. Thus a model including blurring effects was considered and proved to be more accurate than

equation (5.1) [Naderi and Sawchuk, 1978]. Several blurring factors are accumulated in this new model. If we ignore the blurring due to scattering and atmospheric turbulence, and introduce an additional term for quantum fluctuation in this new model, the observation model can be represented as

$$y(i,j) = \sum_{k=i-M}^{i+M} \sum_{l=j-N}^{j+N} [x(k,l) + v_1(k,l) + v_3(k,l)] b(k,l) \quad (5.3)$$

where $v_3(i,j)$ is the quantum fluctuation due to random quanta emission and absorption [Halmshaw, 1982], and $b(i,j)$ is a point-spread function accounting for the blurring factors of chemical diffusion and adjacency effects during the film development, and the blurring factors due to a digitization system. M and N are determined by the width of the unsharpness. To completely solve equation (5.3) for $x(i,j)$, a deconvolution method considering the blurred noise term should be developed. Alternatively, the blurred noise term can be removed first and a straight deconvolution method then used to estimate $x(i,j)$. The latter method is more useful when the processed image has a very low signal-to-noise ratio. The method used in this work to remove the blurred noise term is a model-based adaptive filter. To do this, we collect data into an one-dimensional array for simplicity by raster scanning. Thus, the observation equation becomes

$$y(t) = s(t) + v(t) + n(t) \quad (5.4)$$

where $s(t) = \sum \sum x(k,1)b(k,1)$ and $v(t) = \sum \sum [v_1(k,1)+v_3(k,1)]b(k,1)$. All terms in equation (5.4) have a similar meaning to the terms in equation (5.3) except for the dimensions and an extra term $n(t)$. $n(t)$ is introduced to include effects contributed by dirt on the film and noise of a digitization system [Higgins and Stultz, 1959]. $n(t)$ is Gaussian white noise with a zero mean and a variance of σ_n^2 . A one-dimensional approach is used in this work to simplify the restoration scheme and to increase the execution speed of the program running on a digital computer. A two-dimensional approach may be developed in a future investigation.

Since $v(t)$ is formed by blurring $v_1(i,j)$, $v(t)$ is spatially correlated with a correlation distance larger than the window size of a point-spread function which in turn is larger than the correlation distance of $v_1(i,j)$. Our analysis also shows that the noise fluctuation of an X-ray image is spatially correlated [Zheng and Basart, 1988]. Because $s(t)$ and $v(t)$ are each an autocorrelated process, they can be approximately represented, without loss of generality, as AR(p) and AR(q) models, respectively

$$s(t) = \sum_{i=1}^P \phi_i s(t-i) + w_1(t) \quad (5.5)$$

and

$$v(t) = \sum_{i=1}^q \theta_i v(t-i) + w_2(t) \quad (5.6)$$

where $w_1(t)$ is $iid(0, \sigma_{w1}^2)$ and $w_2(t)$ is $iid(0, \sigma_{w2}^2)$. The mean of $s(t)$ should be removed before modeling; otherwise, there is a constant term in equation (5.5). equations (5.5) and (5.6) show that $s(t)$ and $v(t)$ can be related to their weighted neighboring points by ϕ_i and θ_i plus uncorrelated normally-distributed residuals. AR modeling theory can be found in many textbooks [Fuller, 1976; Abraham and Ledolter, 1983; Graupe, 1984; Box and Jenkins, 1976; Pankratz, 1983].

Coefficients ϕ_i and θ_i can be found by the autocovariance of $s(t)$ and $v(t)$. For instance, p linear equations can be formed by finding the autocovariance of equation (5.5) from lag 1 to lag p (assuming that the mean of $s(t)$ is zero)

$$E[s(t)s(t-h)] = \sum_{i=1}^p \phi_i E[s(t-i)s(t-h)] \quad 1 \leq h \leq p \quad (5.7)$$

which is

$$C_s(h) = \sum_{i=1}^p \phi_i C_s(|i-h|) \quad 1 \leq h \leq p \quad (5.8)$$

where $E[\dots]$ is the mathematical expectation operator and $C_s(h)$ is the autocovariance of $s(t)$ at lag h since $E[s(t)]=0$. equation (5.8) can be written in a matrix form as

$$\begin{bmatrix} C_s(0) & C_s(1) & \dots & C_s(p-1) \\ C_s(1) & C_s(2) & \dots & C_s(p-2) \\ \vdots & \vdots & & \vdots \\ C_s(p-1) & \cdot & \dots & C_s(0) \end{bmatrix} \begin{bmatrix} \phi_1 \\ \phi_2 \\ \vdots \\ \phi_p \end{bmatrix} = \begin{bmatrix} C_s(1) \\ C_s(2) \\ \vdots \\ C_s(p) \end{bmatrix} \quad (5.9)$$

equation (5.9) is called the Yule-Walker equation [Box and Jenkins, 1976] that can be iteratively solved for ϕ_i [Levinson, 1947; Jayant and Noll, 1984]. The coefficients θ_i can be obtained by a similar method.

A process modeled by an AR model should be at least wide-sense stationary [Fuller, 1976]. An $N \times N$ running window, which is much smaller than the size of the image, is used in the raster scanning to obtain an approximately wide sense stationary sample sequence. For each running window, autocovariances of $s(t)$ and $v(t)$ have to be found before the Yule-Walker equation is solved. The autocovariance of equation (5.4) is

$$C_Y(h) = C_S(h) + C_V(h) + \sigma_n^2 \delta(h) \quad 0 \leq h \quad (5.10)$$

where $C_Y(h)$ and $C_V(h)$ are autocovariances of $y(t)$ and $v(t)$, respectively, and $\delta(h)$ is a delta sequence. The cross terms $C_{sn}(h)$ and $C_{vn}(h)$ are zero due to $E[s(t)n(t-h)] = 0$ and $E[v(t)n(t-h)] = 0$. The cross terms $C_{sv}(h)$ is zero due to $E[s(t)v_2(t-h)] = 0$ and $E[s(t)v_3(t)] = 0$. $C_Y(h)$ is calculated from observations, and σ_n^2 is specified for a certain degree of dirt on the film and the quality of a specific digitization system. For simple and practical proposes, in our method $C_V(h)$ is first estimated (call it C_{v0}) in a flat area of an image, and is then calibrated for each running window by an empirical formula:

$$C_V(h) = k_1(t) C_{v0}(h) C_Y(h) / (C_{ym} k_2) \quad (5.11)$$

where $k_1(t)$ is a nonlinear term related to the optical density, C_{ym} is the variance of an entire image, and k_2 is a constant which is higher than 1.0 and related to signal-to-noise ratio. $k_1(t)$ varies from 0.8 to 1.2, corresponding to an optical density region of 1.5 to 3.5, respectively. Finally, knowing $C_v(h)$, $C_y(h)$, and σ_n^2 , $C_s(h)$ can be obtained by equation (5.10). Every estimated variance, where $h=0$, must be semi-positive definite [Fuller, 1976]; if not, it should be forced to be semi-positive definite.

IMAGE ESTIMATION BY ADAPTIVE KALMAN FILTERING

An adaptive Kalman filtering technique is applied to estimate $s(t)$ with known $y(t)$ in equation (5.4). A Kalman filter can separate two or more stochastic processes whose spectra overlap. The separation is done by the least-squares error criterion. The conditions for applying a Kalman filter for such separation are that the first and the second order statistics of each stochastic process are known and the processes are linearly additively mixed. The known factors for this condition can also be power spectral forms or stochastic difference equations (AR models) of the processes. Kalman filter theory is introduced in many sources [Kalman, 1960; Kalman and Bucy, 1961; Brown, 1983; Gelb, 1974]. Applications of Kalman filtering for image restoration can be found in many sources [Woods and Radewan, 1977; Suresh and Shenoi, 1981; Rajala and de Figueiredo, 1981; Dikshit, 1982; Biemond et al., 1983; Tekalp et al., 1986].

Since an entire image is generally a nonstationary process, the above modeling procedure is based on an $N \times N$ running window in which the process is approximately wide sense stationary. Thus, parameters of a Kalman filter vary from window-to-window, and the Kalman filter is adapted to local features of an image. For each window, the p states are denoted as $\{x_i; i=1,2, \dots, p\}$ for $s(t)$ and q states as $\{x_{p+j}(t); j=1,2,\dots,q\}$ for $v(t)$. Then, system equations (5.5) and (5.6) can be written as

$$\begin{bmatrix} x_1(t) \\ x_2(t) \\ \vdots \\ x_p(t) \\ x_{p+1}(t) \\ x_{p+1}(t) \\ \vdots \\ x_{p+q}(t) \end{bmatrix} = \begin{bmatrix} \phi_1 & & & & & & & \\ \phi_2 & & & & & & & \\ \vdots & & & & & & & \\ \vdots & & & & & & & \\ \phi_p & 0 & \dots & 0 & & & & \\ \hline & & & & \theta_1 & & & \\ & & & & \theta_2 & & & \\ & & & & \vdots & & & \\ & & & & \vdots & & & \\ & & & & \theta_q & 0 & \dots & 0 \end{bmatrix} \begin{bmatrix} x_1(t-1) \\ x_2(t-1) \\ \vdots \\ x_p(t-1) \\ x_{p+1}(t-1) \\ x_{p+2}(t-1) \\ \vdots \\ x_{p+q}(t-1) \end{bmatrix} + \begin{bmatrix} w_1(t) \\ 0 \\ \vdots \\ 0 \\ w_2(t) \\ 0 \\ \vdots \\ 0 \end{bmatrix} \quad (5.12)$$

and the observation equation (5.4) can be represented as

$$y(t) = [1 \ 0 \ \dots \ 0 \ 1 \ 0 \ \dots \ 0] [x_1(t) \ \dots \ x_{p+1}(t) \ \dots \ x_{p+q}(t)]^T + n(t) \quad (5.13)$$

where $O(q,p)$ and $O(p,q)$ are $q \times p$ and $p \times q$ zero matrices, respectively. $I(p-1)$ and $I(q-1)$ are $(p-1) \times (p-1)$ and $(q-1) \times (q-1)$ identity matrices, respectively. It can be shown that $x_1(t)=s(t)$ and $x_{p+1}(t)=v(t)$ and equation (12) are equivalent to equations (5.5) and (5.6), and that equation (5.13) is equivalent to equation (5.4). equations (5.12) and (5.13) can also be written as

$$\mathbf{x}(t) = \Phi(t)\mathbf{x}(t-1) + \mathbf{w}(t) \quad (5.14)$$

and

$$y(t) = H(t)\mathbf{x}(t) + n(t) \quad (5.15)$$

where $\mathbf{x}(t)$ is a $((p+q) \times 1)$ process state vector at lag t , $\phi(t)$ is a $((p+q) \times (p+q))$ state transition matrix, $\mathbf{w}(t)$ is a $((p+q) \times 1)$ white driving vector, and $\mathbf{H}(t) = \{h_i; i=1,2,\dots,p+q\}$ in which $h_i(t)=0$ except for $h_1(t)=1$ and $h_{p+1}(t)=1$. The covariance matrix for $\mathbf{w}(t)$ is denoted by $\mathbf{Q}(t) = \{Q_{ij}(t); 1 \leq i \leq p+q, 1 \leq j \leq p+q\}$ in which $Q_{ij}(t)=0$ except for $Q_{11}(t) = \sigma_{w1}^2$ and $Q_{(p+1)(p+1)} = \sigma_{w2}^2$. The variance of $n(t)$ is denoted by R and $R = \sigma_n^2$.

Assuming that an estimated state vector is denoted by $\mathbf{x}(t|t)$, a Kalman filter estimates $\mathbf{x}(t|t)$ by minimizing the individual terms along the major diagonal of the estimation error covariance matrix

$$\mathbf{P}(t|t) = E[(\mathbf{x}(t) - \mathbf{x}(t|t))(\mathbf{x}(t) - \mathbf{x}(t|t))^T] \quad (5.16)$$

where the elements along the major diagonal represent the estimation error variances for the states. With the above definitions and known quantities, a Kalman filter estimates $\mathbf{x}(t)$ by the following recursive procedure [Brown, 1983]:

1. Enter the recursive loop with an initial a priori state vector $\mathbf{x}(t|t-1)$ and an a priori error covariance matrix $\mathbf{P}(t|t-1)$.
2. Compute a Kalman gain vector

$$\mathbf{K}(t|t) = \mathbf{P}(t|t-1)\mathbf{H}^T[\mathbf{H}\mathbf{P}(t|t-1)\mathbf{H}^T + R]^{-1} \quad (5.17)$$

3. Estimate a state vector

$$\mathbf{x}(t|t) = \mathbf{x}(t|t-1) + \mathbf{K}(t|t)[y(t) - \mathbf{H}\mathbf{x}(t|t-1)] \quad (5.18)$$

4. Evaluate the error covariance matrix

$$\mathbf{P}(t|t) = [\mathbf{I} - \mathbf{K}(t|t)\mathbf{H}]\mathbf{P}(t|t-1) \quad (5.19)$$

5. Predict the next state vector and the error covariance matrix

$$\mathbf{x}(t+1|t) = \Phi(t)\mathbf{x}(t|t) \quad (5.20)$$

$$\mathbf{P}(t+1|t) = \Phi(t)\mathbf{P}(t|t)\Phi^T(t) + \mathbf{Q}(t) \quad (5.21)$$

This process is iterated from the first sample to the last sample in a running window from step 2 to step 5. Then, this procedure is repeated for another running window until all samples of an image are processed. The recorded $\mathbf{x}_1(t|t)$ is an estimation of $\mathbf{s}(t)$ which forms an enhanced image. The recorded $\mathbf{x}_{p+1}(t|t)$ is an estimation of $\mathbf{v}(t)$ which forms a noise image. In this case, noise $\mathbf{v}(t)$ is treated as a signal process and the enhancement is done by optimally separating the two processes.

ENHANCING INDUSTRIAL RADIOGRAPHIC X-RAY IMAGES

The above modeling and filtering techniques have been applied to enhance low-contrast noisy industrial radiographic X-ray images. The image shown in Figure 5.1(a) is digitized from a radiograph of a weld from the Martin Marietta Cooperation. The digitization system represents an image with an 8-bit intensity resolution and a 480 x 512 frame size. A subimage of size 256 x 256 representing a physical area of $8.2 \times 8.2 \text{ mm}^2$ was used for enhancement. The image shown in Figure 5.1(a) has low contrast with the flaws barely visible. The maximum flaw size in the images is about $1.5 \times 0.5 \text{ mm}^2$. The background trend was removed from the images [Doering and Basart, 1988]. If we define a signal-to-noise ratio as

$$S/N = \frac{|\text{peak of flaw} - \text{mean of an image}|}{\text{standard deviation of noise}}, \quad (5.22)$$

the S/N of Figure 5.1(a) is about 2.8. A slice plot of Figure 5.1(a) from point (108,1) to point (108,256) is shown in Figure 5.1(b) where (1,1) is at the lower left corner of the image. $C_{v0}(0)$ and $C_{v0}(1)$ were measured from the top right corner in a box (220,200)x(250,250), to be 29.2 and 14.1, respectively. C_{ym} is about 53.3 and k_2 is 1.2. AR(1) processes were used to describe $s(t)$ and $v(t)$. σ_n^2 was assumed as 0.6. ϕ_1 , θ_1 , $C_v(h)$, $C_s(h)$, and $Q(t)$ were automatically computed for each running window. Then, a bank of second order Kalman filters was used to estimate $s(t)$. The size of the running window was 8x8 with the

center 4x4 pixels among processed pixels were saved as the estimated $s(t)$. The output of the Kalman filter was normalized to a range of 256 gray levels. The image of the estimated $s(t)$ is shown in Figure 5.2(a) and the slice plot from point (108,1) to point (108,256) is shown in Figure 5.2(b) which can be compared to Figure 5.1(b). The contrast in Figure 5.2(a) is noticeably higher than that of Figure 5.1a. Since the Kalman filter enhances an image by separating processes, which is conceptually different than a smoothing method, the background noise of Figure 5.2(a) is still present but with a remarkably-reduced fluctuation level. This can also be verified by the comparison between Figure 5.1(b) and Figure 5.2(b). The S/N of Figure 5.2(a) is about 7.2. From the enhanced image, one can easily find flaws, and possibly define sizes of the flaws if the blurring is not severe. Many image analysis methods such as edge detection and feature extraction can then be applied to enhanced images [Zheng and Basart, 1988].

The algorithm was implemented on a VAX 11/780 computer with FORTRAN. Because a Kalman filter has a recursive property, it can be relatively fast if the order of the Kalman filter is not too high. Without counting I/O time, the CPU time used to filter a 256x256 image was about two minutes which is comparable to that of many smoothing techniques.

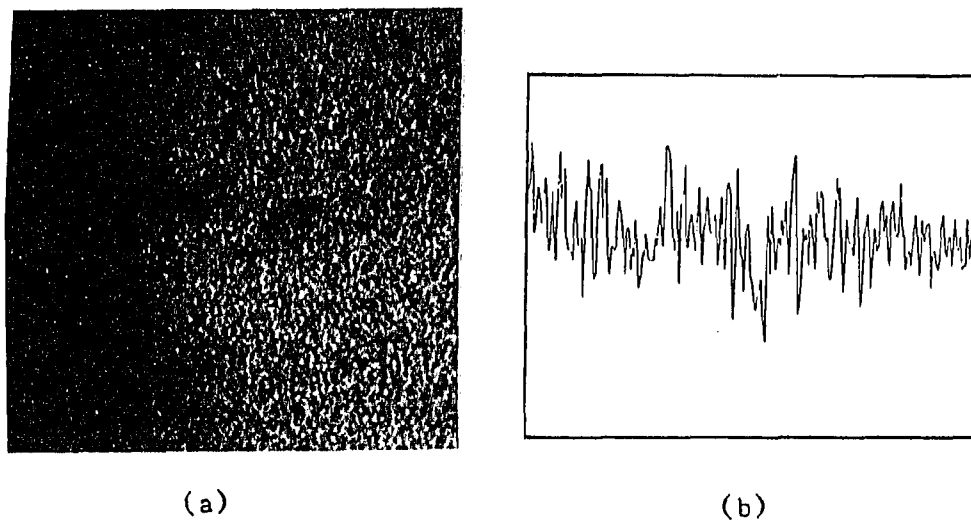


Figure 5.1. An X-ray radiograph of a weld from Martin Marietta Corporation. (a) A digitized NDE radiographic X-ray image, (b) A slice through the plot in (a) from point (108,1) to point (108,256)

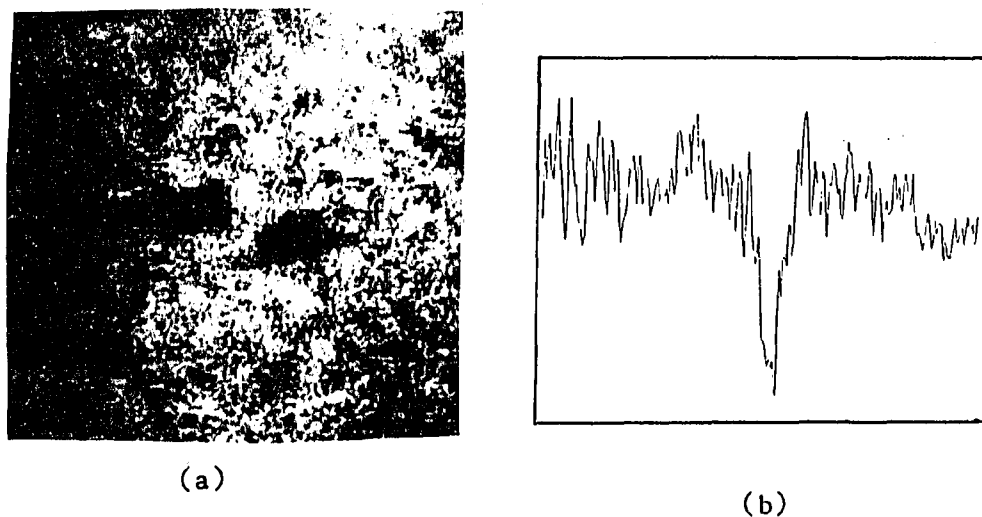


Figure 5.2. An image enhanced by the Kalman filtering. (a) An enhanced image from Figure 5.1(a), (b) A slice through the plot in (a) from point (108,1) to point (108,256)

ACKNOWLEDGMENT

We thank James Lloyd and David Fischer of Martin Marietta for providing radiographs, and thank Joseph Gray for many helpful discussions. This work is supported by the NSF University/Industry Center for NDE at Iowa State University.

REFERENCES

- Abraham, B., and J. Ledolter. 1983. Statistical Methods for Forecasting. John Wiley and Sons, New York.
- Basart, J. P., Y. Zheng, and E. R. Doering. 1987. Application of adaptive region Kalman filtering to x-ray images in NDE. Pages 767-772 in D. O. Thompson, and D. E. Chimenti, ed. Review of Progress in Quantitative NDE. Vol. 6A. Plenum Press, New York.
- Biemond, J., J. Rieske, and J. J. Gerbrans. 1983. A fast Kalman filter for images degraded by both blur and noise. IEEE Trans. Acoust., Speech, Signal Processing ASSP-31:1248-1256.
- Box, G. E. P., and G. M. Jenkins. 1976. Time series analysis: forecasting and Control. Holden-Day, San Francisco, CA.
- Brown, R. G. 1983. Introduction to random signal analysis and Kalman filtering. John Wiley & Sons, New York.
- Dainty, J. C., and R. Shaw. 1974. Image Science. Academic Press, New York.
- Dikshit, S. S. 1982. A recursive Kalman window approach to image restoration. IEEE Trans. Acoust., Speech, Signal Processing ASSP-30:125-140.
- Doering, E., and J. P. Basart. 1988. Application of polynomial regression to trend removal and detection of void-like structures in industrial x-ray image. In D. O. Thompson, and D. E. Chimenti, ed. Review Quantitative Progress NDE. Vol. 7 (in press). Plenum Press, New York.
- Fuller, W. A. 1976. Introduction to statistical time series. John Wiley and Sons, New York.
- Graupe, D. 1984. Time series analysis, identification and adaptive filtering. Robert E. Krieger Publishing Company, Malabar, Florida.
- Gelb, A., ed. 1974. Applied optimal estimation. MIT Press, Cambridge, Mass.
- Gray, J. N. 1988. Three dimensional modeling of projection x-ray radiography. In D. O. Thompson, and D. E. Chimenti, ed. Review of Progress in Quantitative NDE. Vol. 7 (in press). Plenum Press, New York.
- Halmshaw, R. 1982. Industrial radiology: theory and practice. Applied Science Publishers Ltd, Englewood, New Jersey.

- Halmshaw, R. 1973. Information theory applied to industrial radiography. Pages 31-64 in R. S. Sharpe, ed. Research techniques in nondestructive testing. Vol. 2. Academic Press, New York.
- Herz, R. H. 1969. The photographic action of ionizing radiations. John Wiley & Sons, New York.
- Higgins, G. C., and K. F. Stultz. 1959. 'Experimental study of rms granularity as a function of scanning spot size. J. Opt. Soc. Am. 49:925-929.
- Huang, T. S. 1966. Some notes on the grain noise in restoration of atmospherically degraded image. Vol. 2:105-109. NSF Summer Study Report, Woods Hole, Mass.
- Jayant, N. S., and P. Noll. 1984. Digital coding of waveforms - principles and applications to speech and video. Prentice-Hall, Englewood Cliffs, New Jersey.
- Kalman, R. E. 1960. A new approach to linear filter and prediction problems. Trans. ASME 82:35-45.
- Kalman, R. E., and R. S. Bucy. 1961. New results in linear filtering and prediction theory. Trans. ASME 83:95-108.
- Kuan, D. T., A. A. Sawchuk, T. C. Strand, and P. Chavel. 1985. Adaptive noise smoothing filter for images with signal-dependent noise. IEEE Trans. Pat. Anal. Mach. Intel. 7:165-177.
- Levinson, N. 1947. The Wiener rms (root mean square) error criterion in filter design and prediction. J. Math. Phys. 25:261-278.
- Marquardt, D. W. An algorithm for least squares estimation of non-linear parameters. 1963. J. Soc. Ind. Appl. Math. 11:431-441.
- Naderi, F., and A. A. Sawchuk. 1978. Estimation of Images Degraded by Film-Grain Noise. Applied Optics 17:1228-1237.
- Pankratz, A. 1983. Forecasting with univariate Box-Jenkins models. John Wiley, New York, NY.
- Rajala, S. A., and R. I. P. de Figueiredo. 1981. Adaptive nonlinear image restoration by a modified Kalman filtering approach. IEEE Trans. Acoust., Speech, Signal Processing ASSP-29:1033-1042.
- Selwyn, E. W. H. 1939. Experiments on the nature of graininess. Photogr. J. 79:513-524.
- Selwyn, E. W. H. 1959. The 12th Renwick Memorial lecture: Scientists' ways of thinking about definition. J. Photogr. Sci. 7:138-147.

- Suresh, B. R., and B. A. Shenoi. 1981. New results in two-dimensional Kalman filtering with application to image restoration. *IEEE Trans. Circuits Syst.* CAS-28:307-319.
- Tekalp, A. M., H. Kaufman, and J. W. Woods. 1986. Identification of image and blur parameters for the restoration of noncausal blurs. *IEEE Trans. Acoust. Speech Signal Processing ASSP-34*: 963-972.
- Woods, J. W., and C. H. Radewan. 1977. Kalman filtering in two dimensions. *IEEE Trans. Inform. Theory* IT-23:473-482.
- Zheng, Y., and J. P. Basart. 1988. NDE x-ray image modeling and adaptive filtering considering correlated noise. In D. O. Thompson, and D. E. Chimenti, ed. *Review of Progress in Quantitative NDE*. Vol. 7 (in press). Plenum Press, New York.

PART VI. KALMAN FILTERING VLA PHASE DATA WITH A SUPERCOMPUTER

PROBLEM DEFINITION

The earth's atmosphere plays a significant role in both optical and radio astronomy. The precision with which the position and sizes of radio sources can be determined is ultimately limited not only by the size of the telescopes but also by atmospheric effects. As is well known, the radio interferometer and antenna array synthesis techniques, such as the VLA (Very Large Array) [Napier et al., 1983], have greatly improved the resolution of radio source observations. Resolution increases as the baseline (separation between antennas) of the interferometer increases. An interferometer or correlation array measures the Fourier transform form of a radio source's brightness distribution. This transform is called the complex visibility. The phase of the complex visibility carries very important information about the structure and location of the source. Unfortunately, the phase is distorted due to many factors. The most important of these is the earth's atmosphere; the nonhomogeneity of the refractive index of the atmosphere causes the phase to fluctuate in an analytically unpredictable manner. This phase distortion smears the radio image and limits an instrument's performance. Long spatial baselines are required to obtain high resolutions, but as the geometrical baseline and the sky frequency increase, the phase fluctuation also increases [Armstrong and Sramek, 1982]. In the worst case, the phase information of the visibility data becomes useless.

STANDARD CORRECTION TECHNIQUES

Because of the stochastic property of the phase, restoration of the phase becomes very difficult. However, if we observe a point source, the phase fluctuation can be found. Employing this idea, empirical calibration [Hjellming, 1982] and self-calibration [Schwab, 1980] techniques have been developed. Empirical calibration provides the long-run correction, and self-calibration offers short time scale correction. The realizations of these two techniques are based on various assumptions. The assumption for the calibration is that the phase fluctuation is close to being wide-sense stationary, and the assumption for the self-calibration is that the final image (a deconvolution result [Hogbom, 1974]) contains one or more point-like sources or good models. Practice has shown that self-calibration works efficiently to improve the phase data for moderate amounts of atmospheric variation. But if the phase fluctuation is severe, the calibration cannot provide good initial corrections and the deconvolution cannot provide good point-like components; as a result, self-calibration cannot perform well and the image will still be fuzzy.

ARIMA MODELING AND KALMAN FILTERING METHOD

We are developing a new technique to reduce the phase fluctuation. Our approach combines ARIMA modeling theory [Box and Jenkins, 1976; Pankratz, 1983], recursive Kalman filtering theory [Kalman, 1960; Brown, 1983], and existing astronomical data processing algorithms. There are several reasons for taking this approach:

1. The stochastic property of the phase can be incorporated.
2. The phase is highly correlated in time order.
3. ARIMA models are optimal models for stochastic processes and the Kalman filters are optimal linear filters for random signals.
4. The ARIMA modeling procedure and the Kalman filtering processing can be realized with a digital computer.

The correlated components in the data can be extracted by reasonable mathematical representations of the data. The remaining components are considered to be the uncorrelated components. Knowledge of these two components is necessary for a stochastic filtering process. In order to obtain these two components, we have applied the Box-Jenkins modeling method to the VLA phase data. The modeling outputs are ARIMA processes which represent VLA phases. Then the spectral form of an ARIMA model can be estimated so that conventional spectral analysis and signal processing algorithms can be used. In addition, an ARIMA representation can be easily transformed to state-space form so Kalman filters can be utilized for the phase correction.

Our preliminary testing and simulation work show that the phase error can be reduced by Kalman filtering, resulting in an improved image.

COMPUTATIONAL TIME PROBLEM

There is a computational time problem in applying our modeling and filtering method to the VLA data. In the VLA, there are 4×351 channels to produce the complex visibility data and there are several hundred averaged samples to be generated in each channel. In order to process these data, the following routines are necessary:

1. Recursive Kalman filter routine. This includes a series of matrix operations such as multiplication and addition. The orders of matrices depend upon the order of the models.
2. ARIMA model identification and estimation routines. They include autocorrelation function, cross-correlation function, and partial autocorrelation function estimations. They also include non-linear least squares estimation, prediction, and many statistical tests such as t-test, chi-squared, and normal goodness test.
3. Data base manipulation and data pre-processing routines. They include data fetching and storing, and a variety of data transformations and segmentation.

For each data point, the filter routine executes 10 matrix or vector-matrix multiplications, 5 matrix additions or subtractions and one matrix inversion. Suppose we only consider one measurement at a time, then the matrix inversion and half of the matrix or vector-matrix multiplications become scalar operations. Assuming that we have a data set which contains 3 hours of observations, with each sample being an

average over 20 seconds, we will then have 758,160 samples for the 4×351 channels. To filter these data, we have to finish more than 2.2 million matrix multiplications, about 1.5 million vector-matrix multiplications, about 1.5 million matrix additions, and about 3.8 million scalar multiplications. If the order of the matrices, which is the order of a Kalman filter, is chosen as 4, the execution time (CPU time) for filtering these data will be about two and one-half hours in the VLA DEC-10 computer. If we use adaptive Kalman filters, which may be necessary, the number mentioned above will be increased by many times. In addition, the model identification and estimation routines contain thousands of computations of autocorrelation functions and partial-autocorrelation functions, and include more than 5.6 thousand parameter estimations based on a non-linear least squares method. Because a large amount of CPU time is required and there are many repetitive programs to processing a large amount of data in the VLA, this modeling and filtering processing is limited in the present VLA computer system. Under the aid of a supercomputer, this problem can be solved.

OPTIMIZATION AND VECTORIZATION OF THE KALMAN FILTERING ROUTINE ON THE
BCS CRAY X-MP/24 COMPUTER

For experimental purposes, we ran the Kalman filtering routine with a small data set, which contains 44,850 samples, in both the VLA DEC-10 computer and the BCS (Boeing Computer Services) CRAY X-MP/24 computer. The DEC-10 has 768 kilo-words of main memory, a 36-bit word length, and 2.5 MIPS. The CRAY X-MP/24 has 4 million words of main memory, a word length of 64 bits, a 9.5 ns CPU cycle time and a peak vector operation rate of 200 MFLOPS (78 MFLOPS for the BCS CRAY X-MP [Kenneth, 1985]). The job mentioned above took 323.27 seconds of CPU time in the DEC-10 computer while it took only 29.69 seconds in the BCS CRAY X-MP/24. (If we access the data base in a normal manner, this job will take 544.96 seconds of DEC-10 CPU time). The speed-up factor is 10.9. After we applied the optimization and vectorization techniques, the running time in the latter machine was reduced to 9.85 seconds, providing a speed-up factor of 32.8. The speed-up factor of the optimization and vectorization is about 3. Of the total execution time on the CRAY X-MP, the pure computational time (without counting the I/O operation time) was reduced from 22.79 seconds to 2.95 seconds by the optimization and vectorization.

A supercomputer like the CRAY X-MP/24 usually has a vector processing architecture such that its high performance is largely dependent upon the structure of the program run on the machine. It is a waste of resources to use a vector processor as a scalar machine.

A general phenomenon for many data processing programs is that a high percentage of the total computation time may be spent in several DO loops (with only a few statements) no matter how large and how sophisticated a program may be. Finding these DO loops and performing optimization and vectorization of them is crucial in obtaining efficient program operation [Boeing Computer Services, 1984]. We have employed the following three major strategies to optimize and vectorize our filtering program:

1. Optimization with the application software and computational kernels. Since the computational kernels are written in assembly language, they have higher execution speed than those written in high-level languages. We used HSMMPs and SDOT subroutines of the BCS's Vectorpack [Boeing Computer Services, 1985] to replace the original matrix and vector-matrix multiplication FORTRAN source codes. The computation time of the associated codes was reduced from 13.0 seconds (53% of the total execution time) to 2.5 seconds (only 13% of the overall execution time).
2. Vectorization by restructuring the program. The essential principle of vectorization is variable independence. In order to vectorize our program, we restructured the filter program in both the global and local senses. Global reconstruction means that we changed the basic structure of the entire program and local reconstruction means that we

vectorized each DO loop. The candidate loops for vectorization should not contain I/O, GOTO, IF, and CALL statements, and should have no dependency involving an array or non-linear reference to memory. In this step, we obtained a 3.92 second reduction in run time.

3. Reducing the depth of the subroutines. In order to make clean program structures, some programmers employ a lot of subroutines. However, there are initiation operations required to invoke a subroutine. If a subroutine is called thousands or millions of times, the overhead time used for the initiation cannot be ignored. In the extreme case, a high order depth of subroutines may cause the program to be unexecutable. The subroutine depth of our original routine is 3. The subroutines in the bottom level are called 5 times for processing one sample. The execution time of those subroutines is about 3.0 seconds but the overhead time is 5.437 seconds. We reduced the depth from 3 to 2 by replacing 5 CALL statements with the associated source codes. The revised program became long and the structure is not as clear, but the overhead time was reduced to 0.032 seconds, which is only 0.6% of the original overhead time.

Overall, the results of the benchmark testing is shown in the table below.

Table 6.1. Computer Performance

Performance Improvement: DEC-10 versus BCS CRAY X-MP		
DEC-10	BCS CRAY X-MP	Speedup Factor
323.27 sec.	(*) 29.69 sec.	10.9
323.27 sec.	(**) 9.85 sec.	32.8

*: without vectorization and optimization
 **: with vectorization and optimization

CONCLUSION

The development of our modeling and filtering method is limited by the computation time and memory size. This problem can be solved with the aid of a supercomputer. Our experimental results show that the execution of our filtering program, which was optimized and vectorized, on the BCS CRAY X-MP/24 is 32.8 times faster than that on the VLA DEC-10 computer. This encourages us to develop an adaptive Kalman filtering procedure for the VLA image processing. In order to take advantage of vector processing machines, the program should be vectorized. Those standard routines of the AIPS package at the NRAO such as CALIBRATION, CONVOLUTION, DECONVOLUTION and 2D-FFT are very good candidates for optimization and vectorization on a supercomputer such as a CRAY computer.

REFERENCES

- Armstrong, J. W., and R. A. Sramek. 1982. Observations of tropospheric phase scintillations at 5 GHz on vertical paths. *Radio Science* 17:1579-1586.
- Boeing Computer Services. 1984. Supercomputer vectorization and optimization guide. Engineering Technology Application Division, Boeing Computer Services, Seattle, Wash.
- Boeing Computer Services. 1985. VECTORPACK user's manual. Boeing Computer Services, Seattle, Wash.
- Box, G. E. P., and G. M. Jenkins. 1976. Time series analysis: forecasting and control. 2nd ed. Holden-Day, San Francisco, CA.
- Brown, R. G. 1983. Introduction to random signal analysis and Kalman filtering. John Wiley & Sons, New York.
- Hjellming, R. M. 1982. Calibration. *Proc. NRAO Workshop No.5: Synthesis Mapping*, Socorro, NM.
- Hogbom, J. A. 1974. Aperture synthesis with a non-regular distribution of interferometer baselines. *Astron. Astrophys. Suppl.* 15:417-426.
- Kalman, R. E. 1960. A new approach to linear filter and prediction problems. *Trans. ASME* 82:35-45.
- Kenneth, W. N. 1985. Making the most of supercomputer. *Proc. BCS Supercomputer Summer Institute*, Boeing Computer Services, Seattle, Washington.
- Napier, P. J., A. R. Thompson, and R. D. Ekers. 1983. The very large array: Design and performance of a modern synthesis radio telescope. *Proc. IEEE* 71:1295-1320.
- Pankratz, A. 1983. Forecasting with univariate Box-Jenkins models. John Wiley, New York, NY.
- Schwab, F. R. 1980. Adaptive calibration of radio interferometer data. *Proc. Soc. Photogr. Optic. Instrum. Eng.* 231:18-25.

PART VII. DEBLURRING NDE X-RAY IMAGES AND INFRARED IMAGES BY A MAXIMUM
ENTROPY DECONVOLUTION METHOD

ABSTRACT

A maximum entropy deconvolution method is studied and applied to industrial Non-Destructive Evaluation (NDE) images and infrared (IR) images. One of the factors limiting the quality of these images is unsharpness. Unsharpness is caused by a blurring process which may be modeled by a convolution of an image with a point-spread function (PSF). A practical Maximum Entropy Method (MEM) is developed to deconvolve PSF from images. The MEM solution is obtained by maximizing the entropy of a reconstructed image constrained by the χ^2 statistic which describes the misfit between the restored image and the observed images. A direct convolution is used to find the χ^2 statistic for narrow PSFs and the FFT is used for wide PSFs. If the FFT method is used to evaluate the convolution, two FFTs are required in each iteration. A Newton-Raphson approach is used to find a conditional extremum of the entropy equation subject to conditions enforced by Lagrange multipliers. This MEM is applied to NDE X-ray radiographic images and infrared (IR) images.

INTRODUCTION

NDE image forming mechanisms degrade images by causing unsharpness and noise corruption [Halmshaw, 1973; 1982; Herz, 1969; Dainty and Shaw, 1974; Zheng and Basart, 1988a; 1988b]. Unsharpness blurs edges and details of an image while noise decreases the contrast. There are two main factors causing unsharpness of X-ray radiographic images. One factor arises from the finite size of the radiation source, which is never an ideal point source. It causes an effect analogous to a penumbra in visible light; a sharp edge of a specimen is blurred in an observed image. Another factor causing unsharpness is the scattering of electrons in the film emulsion. The factor causing unsharpness of IR images is the diffuse radiation causing a sharp edge to be fuzzy in the recorded image. Noise degrading the contrast is contributed by random photon emission and absorption processes in a specimen, and random processes in detection devices. These degrading factors will blur edges of an image, obscure details, and reduce contrast. Since information about edges, details, and contrast is important for quantitative NDE work, effective image enhancement and image restoration methods must be developed for the NDE community.

This work introduces a practical Maximum Entropy Method (MEM) for restoring NDE X-ray radiographs and infrared images. MEM has been developed as a powerful tool for both image reconstruction and image restoration. It has been successfully applied in astronomical image processing [Frieden and Swindell, 1976; Wernecke and D'Addario, 1977;

Gull and Daniell, 1978; Skilling et al., 1979; Bryan and Skilling, 1980; Gull and Skilling, 1984; Skilling and Bryan, 1984; Skilling and Gull, 1985; Cornwell and Evans, 1985]. It has been applied in other areas such as radiographs [Burch et al., 1983], and tomography [Minerbo, 1979]. These applications have demonstrated that MEM is a superior technique for producing optimal restorations of images from blurred, incomplete, and noisy data. The restored images have fewer artifacts, better edge definition, and more clear details than images restored by traditional linear methods [Burch et al., 1983].

This work is based on the developments by Burch et al. [1983], and Cornwell and Evans [1985]. Our method uses the entropy measure $-\sum p \log p$ of an image. The misfit between the reconstructed image and the observed image is measured by the χ^2 statistic, which gives a measure of consistency with the observed image. The MEM image is obtained by maximizing the entropy with the condition that χ^2 approaches an expected value enforced by a Lagrange multiplier. An iterative convergence technique (a multivariable Newton-Raphson method) is used for an efficient search finding an extremum of constrained entropy. In this method, an approximation for inverting a large-dimension matrix is made by taking the inverse of the diagonal elements of the matrix [Cornwell and Evans, 1985]. A direct convolution is used to find the χ^2 statistic for a narrow point-spread function (PSF) while an FFT is used to find the χ^2 statistic for a wide PSF. In this method the PSF must be known or must be measurable from an observed image.

Compared to other MEMs this MEM has the advantages of being fast and easy to implement.

THE MAXIMUM ENTROPY METHOD

Let a class of ideal images be characterized by a discrete real-valued random field. A realization of this field is $\mathbf{f} = \{f_i; 0 \leq i \leq N-1\}$ where f_i is an intensity value of a pixel located at i of a specific positive additive image to be reconstructed. N is a total pixel number in an image. The related observational image is denoted by $\mathbf{d} = \{d_i; 0 \leq i \leq N-1\}$. Assuming that the blurring process can be represented by convolving \mathbf{f} with a known spatially invariant PSF \mathbf{h} , the observation equation is given by

$$d_i = \sum_{j=0}^{N-1} f_j h_{i-j} + n_i \quad (7.1)$$

where the subscripts are in one dimension for simplicity and n_i is a noise term with a distribution of $\text{iid}(0, \sigma_i^2)$.

The inverse problem arises when we want to find f_i when given \mathbf{h} and \mathbf{d} . A natural way to approach the solution of this problem is to adjust the sequence $\{f_i\}$ until the reconstructed image \mathbf{f} is consistent with the observed image \mathbf{d} with given constraints. Another problem then rises: how should we adjust f_i ? The MEM provides a unique answer for this problem, which selects f_i in such way as to have much information content of an image and least configuration content as possible. The measure of information content can be defined by the entropy S of a probability distribution [Shannon, 1948]. The entropy of a reconstructed image is

$$S = - \sum_{j=0}^{N-1} p_j \log p_j \quad (7.2)$$

where $p_j = f_j / \sum f$. Thus, a particular probability distribution of f_i is chosen so that f_i fits equation (7.1) by maximizing the entropy S . Then, the solution of the inverse problem consists of selecting a sequence $\{f_i\}$ that has maximum entropy and is consistent with the observational equation under certain constraints. In this case, the reconstructed image is called feasible.

The most simple measure for the difference of (7.1) caused by f_i is the single constraint statistic,

$$\chi^2 = \sum_{k=0}^{N-1} (g_k - d_k)^2 / \sigma_k^2 \quad (7.3)$$

where $g_k = \sum f_j h_{k-j}$. The upper bound χ_u^2 is determined by the values that χ^2 can plausibly take. The condition $\chi^2 \leq \chi_u^2$ defines the set of feasible images which passes the statistical test for consistency. The value of χ_u^2 is about $(N+3.29\sqrt{N})$ for 99% confidence [Skilling and Gull, 1985].

Now, the problem becomes a strict conditional optimization problem: Maximizing S subject to $\chi^2 \leq \chi_u^2$. Thus, as usual for such a problem, we can form a Lagrangian function,

$$Q = S - \lambda \chi^2. \quad (7.4)$$

To reduce the bias of the total intensity Σf_j , a second constraint with another Lagrange multiplier μ is added for the Lagrangian function

$$Q = S - \lambda x^2 - \mu \Sigma f_j \quad (7.5)$$

With a assumption shown late, the maximization of (7.5) is equivalent to maximizing a modified form of entropy [Burch et al., 1983]

$$S = - \Sigma f_j (\log(f_j/A) - 1) \quad (7.6)$$

subject to $x^2 \leq x_u^2$ where

$$A = \exp(\Sigma p_j \log f_j - \mu \Sigma f_j) \quad (7.7)$$

is assumed as a predetermined value. Thus the precise value of A has no effect to the MEM solutions [Burch et al., 1983]. The solution of the conditional extremal equation (7.6) can be found by setting the gradient of Q, $(\nabla Q)_i = \partial Q / \partial f_i$, to zero for a suitable Lagrange multiplier λ

$$f_i = A \exp\{\lambda \partial x^2 / \partial f_i\}. \quad (7.8)$$

Since entropy is intrinsically nonlinear, a nonlinear optimization problem is formed. It is usually solved by an iterative numeric method.

THE COMPUTATIONAL ALGORITHM

The optimization problem is to find a root of

$$\nabla Q(f) = 0 \quad (7.9)$$

where the i -th element of $\nabla Q(f)$ is $\partial Q / \partial f_i$. Since Q is neither linear nor quadratic, an iteration approach must be used to continually refine f_i starting with a flat image $f = \{f_i = A; 0 \leq i \leq N-1\}$ until a feasible image is obtained. There are several numerical iterative methods available to solve equation (7.9) [Skilling and Gull, 1985]. The simplest one is steepest ascent, but it is known to be inefficient. The conjugate gradient technique can be used to improve steepest ascent algorithm, but it requires Q to be quadratic. A method using three search directions was shown by Burch et al. [1983] to be powerful, however, 15-20 iterations with 12 two-dimensional FFTs for each iteration are needed. Cornwell and Evans [1985] proposed a simple Newton-Raphson method in their MEM for radio source image reconstruction, and, in practice, showed it to be effective and fast. It requires two FFTs for each iteration. We used a Newton-Raphson method in our MEM even though the nature of the objective function (7.4) of the optimization and the domains of the observational data are different than the Cornwell-Evans' method. References for the multivariable Newton-Raphson method can be found in many text books of numerical-methods [Johnson and Riess, 1982]. In the Newton-Raphson method, the estimation of \hat{f} at step $n+1$ is given by

$$\hat{f}_i^{(n+1)} = \hat{f}_i^{(n)} - (\nabla \nabla Q^{(n)})^{-1} \nabla Q^{(n)} \quad (7.10)$$

where $\hat{f}_i^{(n)}$ is the estimation of f at step n , $\Delta f = \hat{f}_i^{(n+1)} - \hat{f}_i^{(n)}$, and

$$\nabla Q = \nabla S - \lambda \nabla \chi^2 - \mu 1, \quad (7.11)$$

and

$$\nabla \nabla Q = \nabla \nabla S - \lambda \nabla \nabla \chi^2. \quad (7.12)$$

∇S is a vector, and $\nabla \nabla S$ is a purely diagonal matrix. They can be readily found from (7.6). 1 is unit vector. The i -th element of $\nabla \chi^2$ is

$$\partial \chi^2 / \partial f_i = 2 \sum (\sum f_j h_{k-j} - d_k) h_{k-i} / \sigma_k^2 \quad (7.13)$$

which is a convolution between the difference and the PSF. $\nabla \chi^2$ is different than that given by Cornwell and Evans [1985]. The ij -th element of $\nabla \nabla \chi^2$ is

$$(\nabla \nabla \chi^2)_{ij} = 2 \sum h_0 h_{j-i} / \sigma_j^2 \quad (7.14)$$

It reaches a maximum value at $i=j$. If $|j-i|$ is larger than the width of the PSF, $(\nabla \nabla \chi^2)_{ij} = 0$. Since $\nabla \nabla S$ is a purely diagonal matrix, the non-diagonal elements of $\nabla \nabla Q$ are given by $\nabla \nabla \chi$. $(\nabla \nabla Q)^{-1}$ is an inverse matrix of the Jacobian of ∇Q at step n which has a very large dimension of $N \times N$. For simplicity, the nondiagonal elements of the $(\nabla \nabla Q)^{-1}$ were neglected in the Newton-Raphson approach given by Cornwell and Evans [1985]. The justifications for the simplification are: (1) the

sidelobes of h on the nondiagonal elements are small (in our cases, in addition, $h_{|j-i|} = 0$ for large $|j-i|$.) so the inverse Jacobian can be approximated by letting the diagonal elements be the inverse of the diagonal elements of the Jacobian and the nondiagonal elements be zero; (2) $(\nabla\nabla Q)^{-1}$ is a weighting factor for the amount of change related to intensities, and the values of diagonal elements play the same role for weighting. In our case, since the size of the PSF (blurring function) is much smaller than the image size, $\nabla\nabla Q$ is a sparse band matrix filled with many zeros. In addition, $h_0 > h_i$ for all i . Thus, the justifications also hold in our case and the simplification is also used in our method. Therefore, we have approximations of

$$(\nabla\nabla Q)^{-1}_{ii} \sim 1/((\nabla\nabla S)_{ii} - (\nabla\nabla x^2)_{ii}) \quad (7.15)$$

$$(\nabla\nabla Q)^{-1}_{ij} \sim 0 \quad \text{if } i \neq j. \quad (7.16)$$

At this point, all quantities needed for Δf are obtained, and then, Δf can be calculated by equation (7.10).

As iteration progresses, the Lagrange multipliers have to be adjusted to approach their final values. The changes of Lagrange multipliers λ and μ at step n can be estimated by [Cornwell and Evans, 1985]

$$\Delta\lambda = - \Delta x^2 / ||\nabla x^2 \cdot \nabla x^2|| \quad (7.17)$$

$$\Delta\mu = - \Sigma \Delta f_i / N. \quad (7.18)$$

χ^2 should be calculated for (7.17). It can be found by applying Parseval's law to equation (7.3),

$$\chi^2 = \sum_{k=0}^{N-1} (F_k H_k - D_k)^2 / \sigma_k^2 \quad (7.19)$$

where F, H, and D are Fourier transforms of **f**, **h**, and **d**, respectively.

The convergence criterion may be given by

$$||\nabla Q \nabla Q|| < \epsilon ||1 \cdot 1|| \quad (7.20)$$

where ϵ is of order of 0.01 or less. In practice, the convergence criterion may be also given by χ_u^2 , degree of Δf , and the maximum iteration value.

SOME EXAMPLES OF IMAGES RESTORATION

The MEM discussed above was implemented at the Iowa State University Computer Center on the VAX 11/780 computers. Some experiments were done for simulated images and industrial NDE X-ray radiographic and IR images. If the size of a PSF is small (for instance, 5×5), the MEM program is fairly fast because we used a direct convolution. For a wide PSF, the main computation cost is two FFTs per iteration. The number of iterations required is about 15 to 25.

Figure 7.1 is an original computer generated image with a size of 32×32 . The square in the center is 5×5 . Figure 7.2 is a blurred image obtained by convolving Figure 7.1 with a uniform 5×5 blurring matrix. Figure 7.2 has a signal-to-noise ratio of 100 at the brightest part of the image. Figure 7.3 is an MEM image restored from Figure 7.2. The distorted shape and blurred edges are well reconstructed. Since the size of the PSF was small, a direct convolution was used. The number of iterations is 25 for this MEM image. The CPU time is on the order of a few tens of seconds.

Figure 7.4 is a digitized 128×128 industrial NDE IR image from The McDonnell Douglas Aircraft Company. The shape and edges of this image were required to be defined for flaw sizing. However, the shape was blurred due to the difused thermal wave radiation. The edges were too fuzzy to be well defined by conventional edge detection methods. The reconstructed MEM image is given by Figure 7.5. A 15×15 PSF was

used to obtain this MEM image. The number of iterations was 15. Because the width of the PSF is large, FFTs were used to calculate the convolution terms in the MEM. The VAX CPU time used is about 8 minutes. The shape and edges of the MEM image are better defined than the original image.

Figure 7.6 is a digitized 64 x 64 industrial NDE X-ray radiographic image of casting for a valve. The blurred edge was caused by the finite size of the X-ray beam. The MEM image of Figure 7.6 is given by Figure 7.7. A 7 x 7 PSF was used to obtain the MEM image in 15 iterations. The VAX CPU time used was about 2 minutes. The shape and edges of the MEM image are better defined than the original image.

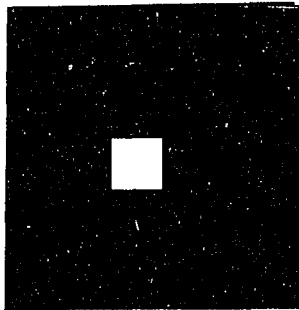


Figure 7.1. An computer generated square image with a size of 32 x 32

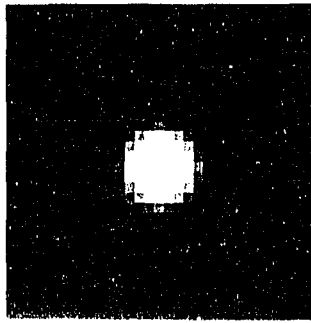


Figure 7.2. A blurred image of Figure 7.1

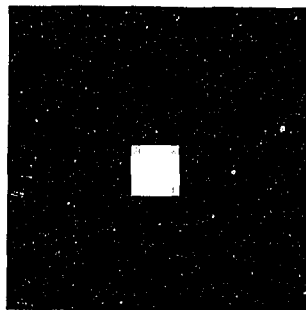


Figure 7.3. An MEM image from Figure 7.2



Figure 7.4. A digitized industrial NDE IR image

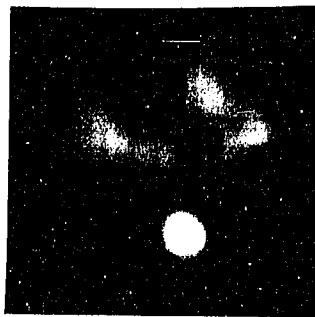


Figure 7.5. A MEM image of Figure 7.4



Figure 7.6. A digitized industrial NDE X-ray radiographic image

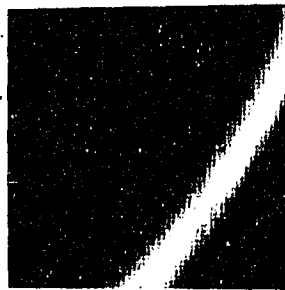


Figure 7.7. An MEM image of Figure 7.6

CONCLUSION

A practical MEM algorithm has been discussed and implemented for industrial NDE images. MEM maximizes image information and minimizes configurational information. It restores a blurred image and gives good definition for shapes, details, and edges of the image. Experiments have shown that this maximum entropy deconvolution method is fast and efficient for practical applications. The main cost of the computational time, which is two FFTs per iteration for large size PSFs and two direct convolutions for narrow PSFs, is less than many other MEMs. However, some parameters of this MEM should be carefully adjusted for a fast and stable convergence. This MEM can be further improved with additional experience.

ACKNOWLEDGMENTS

We thank Daniel C. King of McDonnell Douglas Aircraft Company for providing IR images. This work is supported by the NSF University/Industry Center for NDE at Iowa State University.

REFERENCES

- Bryan, R. K., and J. Skilling. 1980. Deconvolution by maximum Entropy, as illustrated by application to Jet of M87. Mon. Notic. Royal Astron. Soc. 191:69-79.
- Burch, S. F., S. F. Gull, and J. Skilling. 1983. Image restoration by a powerful maximum entropy method. Com. Vision, Graph, Image Proc. 23:113-128.
- Cornwell, T. J., and K. F. Evans. 1985. A simple maximum entropy deconvolution algorithm. Astron. Astrophys. 143:77-83.
- Dainty, J. C., and R. Shaw. 1974. Image science. Academic Press Inc., New York, NY.
- Frieden, B. R., and W. Swindell. 1976. Restored picture of ganymede, moon of jupiter. Science 191:1237-1241.
- Gull, S. F., and G. J. Daniell. 1978. Image reconstruction from incomplete and noise data. Nature 272:686-690.
- Gull, S. F., and J. Skilling. 1984. The maximum entropy method. Pages 267-280 in J. A. Roberts, ed. Indirect imaging. Cambridge University Press, Cambridge, England.
- Halmshaw, R. 1973. Information theory applied to industrial radiography. Pages 31-63 in R. S. Sharpe, ed. Research techniques in nondestructive testing. Academic Press, New York.
- Halmshaw, R. 1982. Industrial radiology: theory and practice. Applied Science Publishers Ltd, Englewood, New Jersey.
- Herz, R. H. 1969. The photographic action of ionizing radiation. John Wiley & Sons Inc, New York.
- Johnson, L. W., and R. D. Riess. 1982. Numerical analysis. Addison-Wesley Publishing Co., Reading, MA.
- Minerbo, G. 1979. MENT: a maximum entropy algorithm for reconstructing a source from projection data. Comp. Graph. Image Proc. 10:46-68.
- Shannon, C. E. 1948. A mathematical theory of communication. Bell System Tech. J., 27:379-423 and 27:623-656.
- Skilling, J., and R. K. Bryan. 1984. Maximum entropy image reconstruction: general algorithm. Mon. Notic. Royal Astr. Soc. 211:111-124.

- Skilling, J., and S. F. Gull. 1985. Algorithms and applications. Pages 83-131 in C. R. Smith and W. T. Grandy, Jr., ed. Maximum-entropy and Bayesian methods in inverse problems. D. Reidel Publ. Co., Boston, MA.
- Skilling, J., A. W. Strong, and K. Bennett. 1979. Maximum-entropy image processing in Gamma-ray astronomy. Mon. Notic. Royal Astr. Soc. 187:145-152.
- Wernecke, S. J., and L. R. D'Addario. 1977. Maximum entropy image reconstruction. IEEE Trans. Computers, C-26:351-364.
- Zheng, Y., and J. P. Basart. 1988a. Image analysis, feature extraction, and various applied enhancement methods for NDE X-ray images. In D. O. Thompson and D. E. Chimenti, ed. Review of Progress in Quantitative NDE. Vol. 7 (in Press). Plenum Press, New York, NY.
- Zheng, Y., and J. P. Basart. 1988b. NDE X-ray image modeling and adaptive filtering considering correlated noise. In D. O. Thompson and D. E. Chimenti, ed. Review of Progress in Quantitative NDE. Vol. 7 (in Press). Plenum Press, New York, NY.

PART VIII. LOCAL FEATURE ENHANCEMENT OF SYNTHETIC RADIO IMAGES BY
ADAPTIVE KALMAN FILTERING

ABSTRACT

An adaptive Kalman filtering technique is developed to improve final radio astronomy images obtained by conventional data processing procedures. This technique focuses on local feature enhancement which is different than other radio astronomy image restoration methods. The enhancement is especially effective in the low S/N region of an image where one may want to distinguish between source structure and background noise. The main steps of this technique are: image segmentation, image modeling, and adaptive Kalman filtering. Compared to traditional image enhancement techniques, this method has advantages of emphasizing local information, estimating a pixel by optimally weighting neighborhoods with a consideration of noise statistics, smoothing noise while preserving edges and source structure, and giving statistics of estimation errors. The technique is applied to enhance VLA (operated by NRAO³) images of radio sources NGC 7027 (planetary nebula) and 3C47 (classical double quasar).

³The National Radio Astronomy Observatory is operated by Associated universities Inc., under contract with the National Science Foundation.

INTRODUCTION

Many signal processing algorithms have been successfully developed to reconstruct and improve synthetic radio images provided by modern interferometric arrays such as the VLA [Thompson et al., 1980; Napier et al., 1983]. Empirical calibration [Hjellming, 1982] and self-calibration [Readhead and Wilkinson, 1978; Schwab, 1980] procedures were designed to correct (u,v) data for atmospheric disturbances. The deconvolution algorithm CLEAN was developed to remove the sidelobes created by Fourier transforming irregularly spaced data and a sparsely filled aperture plane [Hogbom, 1974; 1984; Schwarz, 1978; Clark, 1980; Cornwell, 1983]. The maximum entropy method MEM has been used for efficient deconvolution and reconstruction of extended sources [Frieden and Swindell, 1976; Wernecke and D'Addario, 1977; Gull and Daniell, 1978; Skilling and Gull, 1985; Cornwell and Evans, 1985]. However, these techniques are not perfect and disturbances due to the system and the atmosphere cannot be totally removed from the data, especially for long baselines, high frequencies, and summer weather. The disturbances appear as background noise and blurring of an image. The noise and blurring can make it difficult to distinguish between a real source feature and an artifact caused by noise. The situation becomes worse in a low S/N area. Especially, when intensities of a source and noise are at the same order, one can not confidently define a source structure.

In this work, we develop a scheme to enhance an image so that its contrast can be increased to promote better judgement about a source and the noise. Being different than other radio astronomy image processing methods, our approach enhances local features of an image with local image information. On the contrary, CLEAN and self calibration use global information of an image. To be useful in practice, our method requires little a priori knowledge of an ideal image and the image forming mechanism. We assume that only a distorted observed image is accessible. A model identification method used to find unknown parameters of an image allows the Kalman filtering to be adapted to the local characteristics of an image. Then, the contrast of an image is increased by optimally and recursively estimating each pixel of the image. It is done by optimally weighting neighborhoods of the pixel using spatial noncausality of an image. Because of the nature of adaptive Kalman filter and image segmentation techniques, edges and source structures can be well preserved in a filtered image. Many other noise smoothing methods smooth out the noise, but blur the edges and source structure also. This work is a supplement to radio astronomy data processing methods and can be used in the last stage of a data processing procedure.

There are three main steps in our method. First, low-level image segmentation is used to partition an image into approximately wide-sense stationary regions. Then model identification and parameter estimation methods are applied to each region to find parameters of the

signal and noise models. Finally, an adaptive Kalman filter is used in each region to optimally estimate pixels so that an image is improved. Step one may be skipped when one wants to have fast and efficient processing.

We have used this approach in radio astronomy images and X-ray images in nondestructive evaluation (NDE). Experiments show that our approach is useful for increasing the contrast of an image degraded by both white and colored noise. The procedure is applied to enhance VLA maps of radio sources NGC 7027 (planetary nebula) and 3C47 (classical double quasar).

PRELIMINARY DEFINITIONS

Let an ensemble of two-dimensional views of radio sources be characterized by a discrete finite random field $S=\{S(i,j); 1 \leq i \leq N_1, 1 \leq j \leq N_2\}$ where the image is $N_1 \times N_2$ pixels. Then a view of a specific radio source is a realization of this field and each pixel of its image can be represented by $F_{ij}(S(i,j))$ where F_{ij} is a mapping function describing the image forming mechanism. With interference in observation environments and nonperfect observation instruments, a random noise field $N=[N(i,j)]$ is introduced into the mapping processing. Thus an observed random field $Y=[Y(i,j)]$ can be modeled by

$$Y(i,j) = F_{ij}(S(i,j)) + N(i,j). \quad (8.1)$$

A specific image is then a realization of Y :

$$y(i,j) = F_{ij}(s(i,j)) + n(i,j) \quad (8.2)$$

where the lower case letters indicate realizations of the associated random fields. Assuming that inverse algorithms such as CLEAN and MEM can remove F_{ij} well, we then have

$$y(i,j) = s(i,j) + v(i,j) \quad (8.3)$$

where $v(i,j)$ is a noise term combining $n(i,j)$ and the noise-like disturbance caused by the nonperfect inverse algorithms. The problem is: given observation $y(i,j)$ and statistics of $v(i,j)$, we want to find an estimate $\hat{s}(i,j)$ of $s(i,j)$ which minimizes the summed-squared error

$$E = \sum (s(i,j) - \hat{s}(i,j))^2 \quad (8.4)$$

An approach for finding a solution to this problem is presented in the following three sections.

IMAGE SEGMENTATION BY LOCAL SPATIAL ACTIVITIES

We use a low level image segmentation technique to partition an image into uniform regions which are wide-sense stationary processes. This stationarity is required in the image modeling.

Image segmentation is often used for image classification and image feature extractions. There are many image segmentation techniques such as optimal thresholding [Gonzalez and Wintz, 1977], split and merge [Chen and Pavlidis, 1979], a model based method [Chatterjee and Chellappa, 1987], the Gibbs distribution method [Elliott et al., 1986], region growing and region clustering [Gonzalez and Wintz, 1977], the K-mean method and isodata method [Tou and Gonzalez, 1974], Fuzzy c-mean clustering [Cannon et al., 1986], and a rule-based expert system [Zheng and Basart, 1987]. The key point of all the various methods is that one studies how pixels of an image cluster with respect to some given set of measurements made on the gray levels of these pixels or their neighborhoods. In our method, this is done by clustering local features of an image. The local-feature space in our method consists of local means and masking functions [Rajala and de Figueiredo, 1981] which are related to first and second order statistics, respectively. Then wide-sense stationarity can be obtained by partitioning the local feature space using some thresholds. These thresholds are chosen by examining histograms of local means and masking functions. Since only statistical values are used, and high level knowledge about physical contents of an image is not used in partitioning, this method is called low-level segmentation.

In our method, first and second order statistics of an image are related to the local means and the masking function, respectively. The local mean of an image is defined as

$$m_n(i,j) = \frac{1}{(2n+1)^2} \sum_{p=i-n}^{i+n} \sum_{q=j-n}^{j+n} y(p,q) \quad (8.5)$$

where $y(p,q)$ is the gray level of a pixel at (p,q) . It is a running window average. The window size is $(2n+1) \times (2n+1)$. If the mean value of the noise process is zero and the image process is stationary ergodic, $m_n(i,j)$ will converge to a noise-free mean value at (i,j) on the order of $O_p(2n+1)^{-1}$. Therefore, when n is not too small and the image process is at least locally stationary ergodic, $m_n(i,j)$ is not sensitive to zero-mean noise and is a good estimate of the local mean at (i,j) of a noise-free image. To sort local means, a histogram of local means is calculated and thresholds separating modes of the histogram are used to segment an image. A local minimum-seeking algorithm finds the thresholds by searching slope changes in a smoothed histogram. Local minimum values caused by small fluctuations are ignored. The thresholds are represented by Tl_k where $k = 0, 1, 2, \dots, \beta$; and $Tl_k < Tl_{k+1}$. Then the image is primarily partitioned into subimages Ω_{mk} where $k = 1, 2, \dots, \beta$. A pixel at (i,j) belongs to Ω_{mk} if

$$Tl_{k-1} < m_n(i,j) < Tl_k \quad (8.6)$$

where $Tl_0 = \text{minimum of } m_n(i,j) \text{'s}$ and $Tl_\beta = \text{maximum of } m_n(i,j) \text{'s}$.

The image will be further partitioned according to the second statistic information. This is done by using the masking function. The masking function is defined by

$$M_r(i,j) = \sum_{p=i-r}^{i+r} \sum_{q=j-r}^{j+r} e^{-|| (i,j)-(p,q) ||} \sum_{n=0}^3 D_{pqn} \quad (8.7)$$

where $|| (i,j)-(p,q) ||$ is the Euclidean distance between points (i,j) and (p,q) . D_{pqn} is a slope of a pixel (p,q) in a direction of n [Zheng and Basart, 1988]. The integer r should be greater than the correlation distance of the noise and the order of the AR processes used in the modeling procedure. $M_r(i,j)$ measures spatial activity of an image and is related to the second order statistic. It is also associated with the gradient concept and may be used for edge detection. The standard masking function uses one adjacent pixel on each side of (p,q) to find D_{pqn} and it is very sensitive to noise. By grouping a few pixels at each side of (p,q) , one can reduce noise effects in D_{pqn} . Again, a histogram of the masking function is computed, and the local minimum-seeking algorithm is used to find thresholds separating modes of the histogram. The thresholds are given by $T2_k$ where $k = 0, 1, 2, \dots, a$; and $T2_k < T2_{k+1}$. $T2_k$ s are used to partition the image into subimages Ω_k where $k = 1, 2, \dots, a$. A pixel at (i,j) belongs region Ω_{Mk} if

$$T2_{k-1} < M_r(i,j) < T2_k \quad (8.8)$$

where $T2_0 = \text{minimum of } M_r(i,j)\text{s}$ and $T2_a = \text{maximum of } M_r(i,j)\text{s}$.

Finally, uniform regions Ω_I are obtained by searching for, and identifying, common regions of Ω_{mk} and Ω_{Mk} . The segmentation index $I = 1, 2, \dots, NSEG$; and the maximum number of segments $NSEG \leq \alpha \cdot \beta$. Then each Ω_I is an approximately wide-sense stationary process which is necessary for image modeling and parameter estimation.

SOURCE STRUCTURE AND NOISE DISTURBANCE MODELING

We model an image by specifying its gray level $y(i,j)$ at every point (i,j) of an image. An ensemble of images can be classified as a random field. If this random field is wide-sense markov, its 1D or 2D markovian representation is a recursive process which is often expressed by an autoregressive process with an order of p , $AR(p)$. In other words, if a correlation function can be found for a wide-sense stationary process, and consequently, its unique $AR(p)$ or ARIMA model can be obtained. We reported the application of ARIMA modeling to VLA phase data in Basart and Zheng [1986]. Autoregressive modeling theory and applications can be found in many textbooks [Box and Jenkins, 1976; Pankratz, 1983; Fuller, 1976; Abraham and Ledolter, 1983; Graupe, 1984].

As shown in the last section, we have partitioned an image into wide sense stationary regions. By raster scanning, data of each region are collected into a one dimensional array of which the observation equation (8.3) now becomes

$$y(t) = s(t) + v(t) \quad (8.9)$$

where $y(t)$ is a gray level at a pixel of an observed image, $s(t)$ is an image process to be reconstructed and $v(t)$ is a colored noise sequence (spatially autocorrelated) with a zero mean. Without losing any generality in each wide-sense stationary window, $s(t)$ can be represented as an $AR(p)$ process:

$$s(t) = \sum_{n=1}^p \phi_n s(t-n) + w(t) \quad (8.10)$$

where $w(t)$ is an $iid(0, \sigma_w^2)$. Since $v(t)$ has a zero mean, $E[y(t)] = E[s(t)]$, and it is presumably taken out from $y(t)$. Otherwise, there is a constant term in equation (8.10). In the same window, $v(t)$ may be represented as an AR(q) process:

$$v(t) = \sum_{n=1}^q \theta_n v(t-n) + n(t) \quad (8.11)$$

where $n(t)$ is an $iid(0, \sigma_n^2)$. $v(t)$ can be either colored or white. If $v(t)$ is white, $v(t) = n(t)$ which is a simple form. In practice, noise is always somewhat correlated.

There are many ways to find the coefficients of (8.10) and (8.11). Given $y(t)$ and some parameters, such as q and σ_n^2 or p and σ_w^2 , one may find all unknowns of (8.10) and (8.11) by the model identification method. At the present stage of our work, a simple and straightforward method requiring known, or measurable, statistics of $v(t)$ is developed for modeling. The procedure is given in the following.

Assume that the noise is stationary in an image and the spatial autocovariances of the noise can be measured from a flat region in the image, where the spatial activity is dominated by noise, we then can find θ_n by solving the Yule-Walker equation (Box and Jenkins, 1976):

$$\begin{bmatrix} c_v(0) & c_v(1) & \dots & c_v(q-1) \\ c_v(1) & c_v(2) & \dots & c_v(q-2) \\ \vdots & \vdots & & \vdots \\ c_v(q-1) & \cdot & \dots & c_v(0) \end{bmatrix} \begin{bmatrix} \theta_1 \\ \theta_2 \\ \vdots \\ \theta_q \end{bmatrix} = \begin{bmatrix} c_v(1) \\ c_v(2) \\ \vdots \\ c_v(q) \end{bmatrix} \quad (8.12)$$

where $c_v(h)$ is an autocovariance of $v(t)$ at a lag of h . Equation (8.12) can be iteratively solved for θ_n (Levinson, 1974; Jayant and Noll, 1984). σ_n^2 can be obtained by calculating the variance of

$$n(t) = v(t) - \sum_{n=1}^q \theta_n v(t-n). \quad (8.13)$$

The autocovariance of equation (8.9) is

$$c_y(h) = c_s(h) + 2 c_{sv}(h) + c_v(h) \quad 0 \leq h \quad (8.14)$$

where $c_y(h)$ is the autocovariance of $y(t)$ that can be estimated from the observed image. The cross term c_{sv} should be zero due to $E[s(t)n(t)] = 0$. $c_n(h)$ is the autocovariance of $n(t)$ that can be readily obtained from equation (8.13). In practice, the estimated covariance matrix of $s(t)$ must be semi-positive definite; otherwise, the parameters of the noise should be adjusted. With known $c_s(h)$, $\phi(n)$ can be found with a Yule-Walker equation of order p . σ_w^2 can be estimated by calculating the variance of $w(t)$ with equation (8.10). At this point, we obtain all parameters of the AR models of the signal and

noise processes in equation (8.9) for a running window. This parameter estimation procedure is repeated for each running window until the entire image is processed.

At this point, we have found all unknown parameters of the image system described by equations (8.9) to (8.11). These parameters are needed to separate the noise from the signal by Kalman filtering.

KALMAN FILTERING IMAGES

The Kalman filter is an optimal filter that can separate two or more additive stochastic processes whose spectra overlap. The separation is done with the criterion of least squared error. The Kalman filter is a recursive linear vector filter. Its state space form allows system, signal noise, deterministic, stochastic, and other interested processes be considered and introduced into the filter equations. Kalman filter theory can be found in many references [Kalman, 1960; Kalman and Bucy, 1961; Gelb, 1974; Brown, 1983]. Studies of the Kalman filter in image enhancement can be found in [Woods and Radewan, 1977; Biemond et al., 1983; Tekalp et al., 1986].

The system given by equations (8.9) to (8.11) has to be written in a state-space form and then implemented in the Kalman filter. Let $s(t)=x_1(t)$ and $v(t)=x_{p+1}(t)$, equations (8.10) and (8.11) can be represented by

$$\begin{bmatrix} x_1(t) \\ x_2(t) \\ \vdots \\ x_p(t) \\ x_{p+1}(t) \\ x_{p+1}(t) \\ \vdots \\ x_{p+q}(t) \end{bmatrix} = \begin{bmatrix} \phi_1 & & & & & & & \\ \phi_2 & & & & & & & \\ & \ddots & & & & & & \\ & & I(p-1) & & & & & \\ & & & & 0(p,q) & & & \\ & & & & & & & \\ \phi_p & 0 & \dots & 0 & & & & \\ \hline & & & & 0(q,p) & & & \\ & & & & & \theta_1 & & \\ & & & & & \theta_2 & & \\ & & & & & & I(q-1) & \\ & & & & & & & \vdots \\ & & & & & & & \theta_q & 0 & \dots & 0 \end{bmatrix} \begin{bmatrix} x_1(t-1) \\ x_2(t-1) \\ \vdots \\ x_p(t-1) \\ x_{p+1}(t-1) \\ x_{p+2}(t-1) \\ \vdots \\ x_{p+q}(t-1) \end{bmatrix} + \begin{bmatrix} w_1(t) \\ 0 \\ \vdots \\ 0 \\ w_2(t) \\ 0 \\ \vdots \\ 0 \end{bmatrix} \quad (8.15)$$

where $I(p-1)$ and $I(q-1)$ are $(p-1) \times (p-1)$ and $(q-1) \times (q-1)$ identity matrices. $O(p,q)$ and $O(q,p)$ are $p \times q$ and $q \times p$ zero matrices, respectively. (8.15) can also be written as

$$\mathbf{x}_t = \Phi \mathbf{x}_{t-1} + \mathbf{w}_t. \quad (8.16)$$

Where, by definition, the transition matrix relates \mathbf{x}_{t-1} to \mathbf{x}_t . The observation equation becomes of

$$y(t) = \mathbf{H} \mathbf{x}_t + \epsilon(t) \quad (8.17)$$

where \mathbf{H} is a measurement vector giving a connection between the measurement $y(t)$ and the state vector \mathbf{x}_t . In this case, \mathbf{H} is a $(p+q)$ row vector and $H(i)=0$ (for $i=1,2,\dots,q+q$) except $H(1)=1$ and $H(p+1)=1$. $\epsilon(t)$ is a very small disturbance term. It is introduced in (8.17) for numerical purposes. The variance of $\epsilon(t)$ is denoted by R .

If the data scanning procedure mentioned in the last section is processed with a running window in a segment of an image where the size of the running window is smaller than that of the segment, the estimated parameters of equations (8.10) and (8.11) will be changed not only from segment-to-segment but also from window-to-window. Thus they represent the local nature of an image so that the Kalman filter is adapted to the local features of the image.

The Kalman filter is a set of recursive equations given by the following.

1. Predict the next state vector and the error covariance matrix:

$$\mathbf{x}_{t+1|t} = \Phi \mathbf{x}_{t|t} \quad (8.18)$$

$$P_{t+1|t} = \Phi P_{t|t} \Phi^T + Q \quad (8.19)$$

2. Compute the Kalman gain vector:

$$K_{t|t} = P_{t|t-1} H^T (H P_{t|t-1} H^T + R)^{-1} \quad (8.20)$$

3. Estimate the state vector:

$$\mathbf{x}_{t|t} = \mathbf{x}_{t|t-1} + K_{t|t} (y(t) - H \mathbf{x}_{t|t-1}) \quad (8.21)$$

4. Estimate the error covariance matrix:

$$P_{t|t} = (I - K_{t|t} H) P_{t|t-1} \quad (8.22)$$

where $\mathbf{x}_{t|t}$ is an estimate of \mathbf{x}_t with the (t) th observation, and $\mathbf{x}_{t|t-1}$ is a prediction for \mathbf{x}_t at step $(t-1)$. $P_{t|t}$ is an error covariance matrix which is defined as $E[(\mathbf{x}_t - \mathbf{x}_{t|t})^2]$. Q is a $(p+q) \times (p+q)$ matrix with $Q(1,1) = \sigma_w^2$, $Q(p+1,p+1) = \sigma_n^2$, and all other elements equal to zero.

Given initial values $\mathbf{x}_1|0$ and $P_1|0$ and the first measured data point, the $K_1|1$ and $\mathbf{x}_1|1$ can be found by (8.18) and (8.19), respectively. Thus a recursive filtering procedure begins to separate $s(t)$ (i.e., $x_1(t)$) from the noisy observation $y(t)$. First, a prediction or a best guess of the next state vector can be made by (8.18). Then this guess is corrected by a calibration factor which is given by optimally weighting the difference between an observation and its prediction (8.21). The first element of the calibrated vector is an estimate of $s(t)$. Finally, the P matrix gives a measure of the estimation error by (8.20). This procedure is repeated for the next

observation from step 1 to step 4 until all observations in a window are processed. A forward-backward filter [Brown, 1983] is used in this procedure. For a stationary process, $P_{t|t}$ and $K_{t|t}$ will approach constant values after the first few observations are processed. In order to take advantage of the noncausality of an image, the directions of raster scanning can be chosen from different directions. However, forward and backward filtering and various direction scanning will add considerable extra computational time.

This modeling and filtering procedure is repeated from one window to another and from one segment to another until all pixels of an image are processed. The image consisting of the first component of $\mathbf{x}_{t|t}$ is the desired output.

ENHANCING RADIO SYNTHETIC IMAGES NGC 7027 AND 3C47

The technique introduced above has been used to enhance VLA radio images NGC 7027 (planetary nebula) and 3C47 (classical double quasar). NGC 7027 was observed by J. P. Basart and C. T. Daub using the VLA at a frequency of 14.9649 GHz ($\lambda=2\text{cm}$) with the C configuration [Basart and Daub, 1987]. 3C47 was observed by J. P. Basart and J. O. Burns using the VLA at a frequency of 4.8351 GHz ($\lambda=6\text{cm}$) with the B configuration. The standard VLA data processing procedure was used to edit, calibrate, self-calibrate, map, and CLEAN the data. The resulting maps are shown in Figure 8.1 and Figure 8.2 which have sizes of 92 x 92 and 288 x 288, respectively. These two maps with lower contour levels are given in Figure 8.3 and Figure 8.4. Note that the intensities of the jet in Figure 8.4 are of the same order as background noise. If we define a dynamic range to be a ratio of the peak flux density to the rms value of the background noise, the highest dynamic ranges of the Figure 8.3 and Figure 8.4 are 2000 and 2500, respectively. In the jet area, the dynamic ranges are about 1.0 to 5.5.

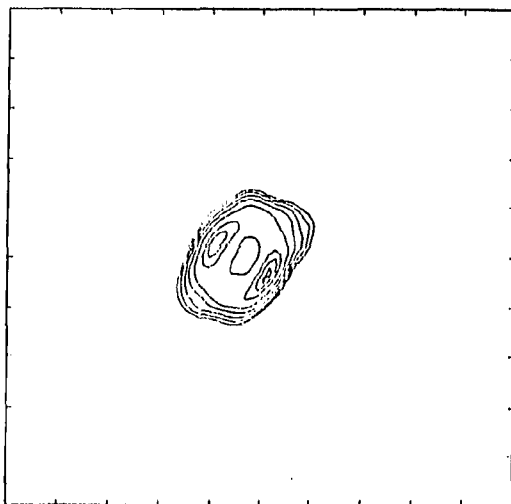


Figure 8.1. A contour map of a radio source NGC 7027. The peak flux density is 0.2616 Jy/beam

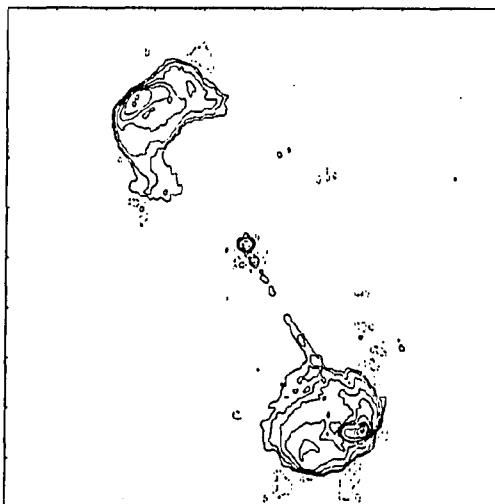


Figure 8.2. A contour map of a radio source 3C47. The peak flux density is 0.1742 Jy/beam. The % contour levels are: -0.2, -0.1, 0.1, 0.2, 0.4, 0.8, 1.6, 3.2, 10, 20, 40, 80, 95

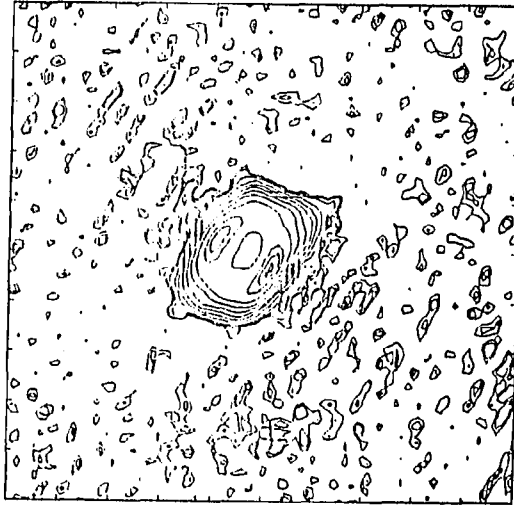


Figure 8.3. A contour map with the additional contour levels 0.04, 0.08, and 0.2 of Figure 8.1

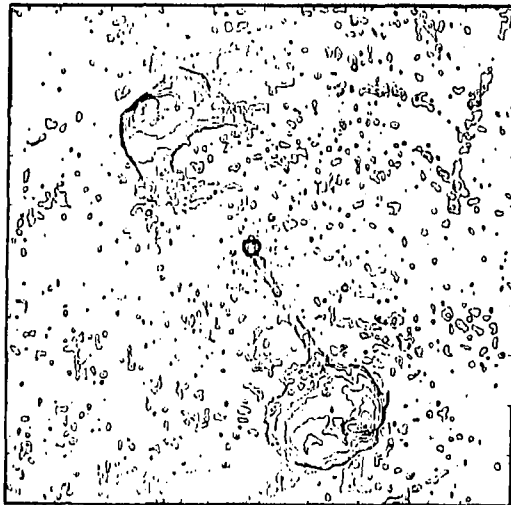


Figure 8.4. A contour map with an additional contour level 0.05 of Figure 8.2

The adaptive Kalman filter with the segmentation technique was used to enhance the NGC 7027 map. The segmentation number NSEG was 5 and the running window size was 8x8. AR(2) models were used for describing the source structure and a white noise model was used for the background noise. The rms of the noise was estimated as 0.129 mJy. The reconstructed map by the Kalman filter is given in Figure 8.5 with the same contour levels as Figure 8.3. The noise level of Figure 8.5 is reduced to 0.048 mJy and the dynamic range was increased to 5400. The estimated error (average rms of estimation error) of the Kalman filtered map is about 0.03 mJy. Since a white noise model was used, the white components of the background noise were well removed but the correlated structure of the disturbance remains in the low-middle area. Those extended structures close to the source edge may be considered as either source extension or source power dispersion due to phase error or nonperfect deconvolution. Since the extended structures have high spatial variances and are highly spatially correlated, they were recognized as source structure by the Kalman filters used and not changed much. It took about 15 minutes of VAX 11/780 CPU time to segment and filter this map. Since there is a search process involved in filtering with the segmentation method, it is relatively computational-time expensive.

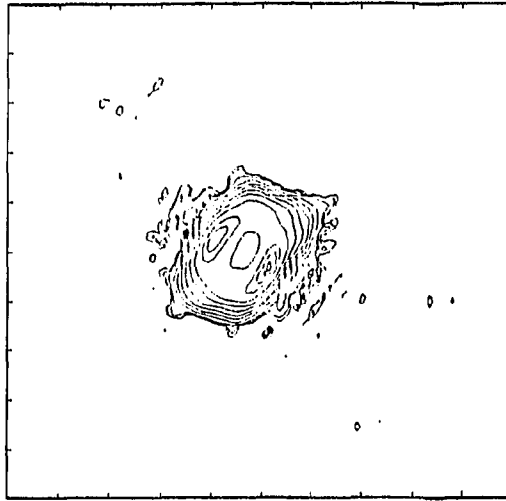


Figure 8.5. The enhanced map from Figure 8.3. The contour levels are the same as Figure 8.3

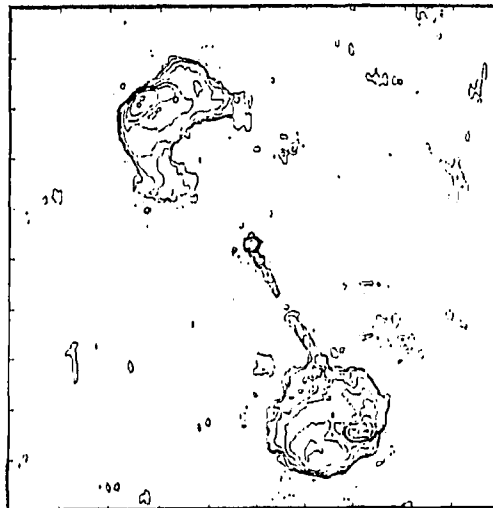


Figure 8.6. An enhanced map from Figure 8.4. The peak flux density and the contour levels are the same as Figure 8.4

The result of Kalman filtering the 3C47 map is shown in Figure 8.6 with the same contour levels of Figure 8.4. Since the map size of Figure 8.4 was 300 x 300, it would take about two and half hours of VAX 11/780 CPU time to filter this map with the same scheme used for NGC 7027. In addition, if we model the noise as a colored process $AR(q)$, the time will be increased by a factor related to $(q+1)$. Therefore, it is not practical. A fast filter should be found. A scheme without segmentation is used for fast filtering. In this scheme, the size of the running window is chosen so that the image process in the window is close to wide-sense stationary for most regions of the image. The price for this scheme is that the filter would not do much for those regions located in the edges of the source. Since the background noise of Figure 8.4 was highly spatially correlated, an $AR(1)$ model was used to model the noise. In this case, the noise was treated as a second signal source. An $AR(1)$ process was used to model the source structure in each window. The low orders of AR models were used for the fast filter. In addition, the program structure was optimized for computational time. The running window size was chosen as 12 x 12. The computation time of this filter scheme without segmentation was two and half minutes of VAX 11/780 CPU time without counting I/O time. The computational time had been reduced so much that this scheme is well suited for practical applications. The rms value of noise in the filtered map is 0.4 mJy while it in the Figure 8.4 is 0.7 mJy. The average rms value of the estimation error for each pixel was 0.25 mJy.

The dynamic range is increased to 4200. The dynamic ranges of the jet structures are about 2.5 to 10.0. The filtered map shows that the background noise was reduced and the jet structure was better exposed. Comparing Figure 8.4 and Figure 8.6, one may have a higher degree of confidence in declaring the discovery or the existence of a jet.

CONCLUSION

A radio astronomy image can be further enhanced by an adaptive Kalman filter. The enhancement is designed for local features. It smooths noise but preserves edges and source structure well. The error covariance matrix of the Kalman filter gives statistics of the estimation error of the filtering which are never given by other image enhancement techniques. This technique is helpful when one wants to identify some source structure in fuzzy regions where intensities are on the same order as the noise. It increases one's degree of confidence when trying to discriminate between source structure and noise. Filtering with a segmentation technique is a complete filtering scheme in theory. However, a filtering scheme without segmentation can be used in practice to reduce computation time. We suggest using this method in the very last stage of image processing. This work may be further developed to enhance radio synthetic maps with partially known patterns of the radio sources.

ACKNOWLEDGMENTS

This work was partially funded by National Science Foundation Grant no. AST-8217135. We thank J. O. Burns and C. T. Daub for sharing VLA images. We thank W. Fuller, R. A. Sramek, T. J. Cornwell, and R. D. Ekers for many helpful discussions.

REFERENCES

- Abraham, B., and J. Ledolter. 1983. Statistical method for forecasting. John Wiley & Sons, Inc., New York, NY.
- Basart, J. P., and Y. Zheng. 1986. Modeling very large array phase data by the Box-Jenkins method. *Radio Science* 21:863-881.
- Basart, J. P., and C. T. Daub. 1987. Temperature and emission-measure distributions for several planetary nebular. *Astrophysical J.* 317:412-422.
- Biemond, J., J. Rieske, and J. J. Gerbrans. 1983. A fast Kalman filter for images degraded by both blur and noise. *IEEE Trans. Acoust., Speech, Signal Processing ASSP-31:1248-1256.*
- Box, G. E. P., and G. M. Jenkins. 1976. Time series analysis: forecasting and control. 2nd ed. Holden-Day, San Francisco, CA.
- Brown, R. G. 1983. Introduction to random signal analysis and Kalman filtering. John Wiley & Sons, New York.
- Cannon, R. L., J. V. Dave, and J. C. Bezdek. 1986. Efficient implementation of the fuzzy c-means clustering algorithms. *IEEE Trans. Pattern Anal. Mach. Intel. PAMI-8:248-255.*
- Chatterjee, S., and R. Chellappa. 1987. Model-based texture segmentation. Pages 235-239. in D. M. Etter, ed. Twentieth Asilomar Conference on Signal, Systems & Computers. IEEE Inc., New York.
- Chen, P. C., and T. Pavlidis. 1979. Segmentation by texture using a co-occurrence matrix and a split-and-merge algorithm. *Comp. Graph. Image Proc.* 10:172-182.
- Clark, B. G. 1980. An efficient implementation of the algorithm "CLEAN,". *Astron. Astrophys.* 89:377-378.
- Cornwell, T. J. 1983. A method of stabilizing the clean algorithm. *Astron. Astrophys.* 121:281-285.
- Cornwell, T. J., and K. F. Evans. 1985. A simple maximum entropy deconvolution algorithm. *Astron. Astrophys.* 143:77-83.
- Dikshit, S. S. 1982. A recursive Kalman window approach to image restoration. *IEEE Trans. Acoust., Speech, Signal Processing ASSP-30:125-140.*

- Elliott, H., H. Derin, and R. Cristi. 1986. Application of the Gibbs distribution to image segmentation. Pages 3-24 in E. J. Wegman and D. J. Depriest, eds. Statistical image processing and graphics. Marcel Dekker, Inc., New York.
- Frieden, B. R., and W. Swindell. 1976. Restored picture of ganymede, moon of jupiter. *Science* 191:1237-1241.
- Fuller, W. A. 1976. Introduction to statistical time series. John Wiley & Sons, Inc. New York.
- Gelb, A., ed. 1974. Applied optimal estimation. MIT Press, Cambridge, MA.
- Gull, S. F., and G. J. Daniell. 1978. Image reconstruction from incomplete and noise data. *Nature* 272:686-690.
- Gonzalez, R. C., and P. Wintz. 1977. Digital Image Processing. Addison-Wesley, Reading, Mass.
- Graupe, D. 1984. Time series analysis, identification and adaptive filtering. Robert E. Krieger Pub. Co., Malabar, Florida.
- Hjellming, R. M. 1982. Calibration. Pages 4.1-4.15 in Proc. NRAO workshop No. 5, Synthesis mapping. NRAO, Socorro, NM.
- Hogbom, J. A. 1974. Aperture synthesis with a nonregular distribution of interferometer baselines. *Astron. Astrophys. Suppl.* 15:417-426.
- Hogbom, J. A. 1984. CLEAN as a pattern recognition procedure. Pages 247-254 in J. A. Roberts, ed. Indirect Imaging. Cambridge University Press, Cambridge, NY.
- Jayant, R. C., and P. Noll. 1984. Digital coding of waveforms: principles and applications to speech and video. Prentice-Hall, Englewood Cliffs, New Jersey.
- Kalman, R. E. 1960. A new approach to linear filter and prediction problems. *Trans. ASME* 82:35-45.
- Kalman, R. E., and R. S. Bucy. 1961. New results in linear filtering and prediction theory. *Trans. ASME* 83:95-108.
- Levinson, N. 1947. The Wiener rms (root mean square) error criterion in filter design and prediction. *J. Math. Phys.* 25:161-278.
- Napier, P. J., A. R. Thompson, and R. D. Ekers. 1983. The very large array: design and performance of a modern synthesis radio telescope. *Proc. IEEE* 71:1295-1320.

- Pankratz, A. 1983. Forecasting with univariate Box-Jenkins models. John Wiley, New York, NY.
- Rajala, S. A., and de Figueiredo, R. I. P. 1981. Adaptive nonlinear image restoration by a modified Kalman filtering approach. IEEE Trans. Acoust., Speech, Signal Processing ASSP-29:1033-1042.
- Readhead, A. C. S., and P. N. Wilkinson. 1978. The mapping of compact radio sources from VLBI data. Astrophys. J. 223:25-36.
- Schwab, F. R. 1980. Adaptive calibration of radio interferometer data. Proc. Soc. Photogr. Optic. Instrum. Eng. 231:18-25.
- Schwarz, U. T. 1978. Mathematical-statistical description of the iterative beam removing technique (method CLEAN). Astron. Astrophys. 65:345-356.
- Skilling, J., and S. F. Gull. 1985. Algorithms and applications. Pages 83-131 in C. R. Smith and W. T. Grandy, Jr., eds. maximum-entropy and Bayesian methods in inverse problems. D. Reidel Publ. Co., Boston, MA.
- Tekalp, A. M., H. Kaufman, and J. W. Woods. 1986. Identification of image and blur parameters for the restoration of noncausal blurs. IEEE Trans. Acoust. Speech Signal Processing ASSP-34:963-972.
- Thompson, A. R., B. G. Clark, C. M. Wade, and P. J. Napier. 1980. The very large array. Astrophys. J. Suppl. 44:151-167.
- Tou, J. T., and R. C. Gonzalez. 1974. Pattern Recognition Principles. Addison-Wesley, Reading, MA.
- Wernecke, S. J., and L. R. D'Addario. 1977. Maximum entropy image reconstruction. IEEE Trans. Computers, C-26:351-364.
- Woods, J. W., and C. Radewan. 1977. Kalman filtering in two dimensions. IEEE Trans. Inf. Theory IT-23:473-482.
- Zheng, Y., and J. P. Basart. 1987. Automatic image segmentation, modeling, and restoration with a rule-based expert system. Pages 421-425 in D. M. Etter, ed. Twentieth Asilomar Conference on Signal, System & Computers. IEEE Inc., New York.

PART IX. T-CALIBRATION: A NEW TECHNIQUE FOR CORRECTING ATMOSPHERIC-
INDUCED PHASE ERRORS OF A SYNTHETIC-APERTURE ANTENNA ARRAY BY TIME
SERIES MODELING AND KALMAN FILTERING

ABSTRACT

A new technique is discussed and developed to correct atmospheric-induced errors in phase data of radio astronomy interferometers and synthetic-aperture antenna arrays. The main feature of this technique is to model and filter the information contents of the phase data in time sequences. T-calibration means correcting astronomical phase data by the time series modeling and Kalman filtering technique. Because the atmospheric phase variations are highly correlated in time, they can be described by stochastic time series models. In conjunction with other radio astronomy data processing algorithms, a time series modeling and parameter estimation technique is developed to obtain noise models and source models from observed phase data. These models can be in the form of stochastic difference equations, auto-correlation, power spectral, or state variable formats ready for further data processing. Once the models are in state variable formats, the Kalman filter is used for optimally extracting source information from noisy data. The resulting synthetic image is then improved by reducing the phase error. The quality of the corrected phase data is quantitatively described by the error covariance matrix of the Kalman filter. This technique has been tested using the VLA (operated by NRAO⁴) and the Hat Creek millimeter interferometer. At the end of this paper, some suggestions for further developments are

⁴The National Radio Astronomy Observatory (NRAO) is operated by Associated Universities, Inc., under contract with the National Science Foundation.

given.

INTRODUCTION

Atmospherically induced phase fluctuation reduces the performance of high-resolution radio astronomy interferometers and synthetic-aperture antenna arrays. Studies have shown that the phase fluctuation is mainly dominated by atmospheric water vapor at centimeter wavelengths and erratically increases as baselines and frequencies increase [Baars, 1967; Basart et al., 1970; Hinder, 1970; Mathur et al., 1970; Wesseling et al., 1974; Hinder and Ryle, 1971; Hargrave and Shaw, 1978; Hamaker, 1978; Dravskikh and Finkelstein, 1979; Han, 1980; Moran and Rosen, 1981; Armstrong and Sramek, 1982; Treuhaft and Lanyi, 1987].

In practice, two calibration methods were developed for reducing the phase error. The first method, called "empirical calibration" using point-like calibrator sources, was designed to correct the long-term phase fluctuation. The second method, named "self-calibration" employing the concept of phase closure and CLEAN components, provides the short-term phase correction [Jennison, 1958; Readhead and Wilkinson, 1978; Schwab, 1980]. The CLEAN method was designed to remove the synthesized beam from a map [Hogbom, 1974; 1984; Schwarz, 1978; Clark, 1980; Cornwell, 1983]. These two calibration methods work well for most observation situations. Exceptions are severe atmospheric conditions, no point-like source in the field-of-view, and no good model of the source.

Another approach to this problem is to use a microwave water vapor radiometer to estimate the atmospheric phase fluctuation. Although this technique has had some success, there is a problem to be solved before the technique can be used routinely [Resch et al., 1984].

Some efforts have also been made for finding statistical models of the phase fluctuation such as the spatial structure function of the phase fluctuations [Tatarskii, 1961; 1971; Ishimaru, 1978; Armstrong and Sramek, 1982; Treuhaft and Lanyi, 1987]. With a different approach to the problem than the structure function, a time series modeling technique was proposed to describe the stochastic properties of the phase fluctuation in time [Basart and Zheng, 1986]. It was shown that time series models provide good short-term forecasting for VLA phase data and reasonably describe the behavior of the atmospheric phase variation for certain time durations. We also reported that the phase fluctuation is highly correlated in time. Thus, it was natural to search a new avenue to restore phase information in time order by which the information of phase fluctuation in time is used.

In this work, a time series modeling and Kalman filtering technique is introduced to correct the phase fluctuation in time. We expect this alternative data correction technique to be used to improve the performance of high-resolution radio astronomy interferometers and synthetic-aperture antenna arrays when the performance of standard calibration methods are not good enough or when the number of interferometers of an array is not enough to apply self-calibration.

The Kalman filter is a recursive and optimal linear filter couched in state-variable theory for modern control systems. The main feature of the Kalman filter is that the filter can optimally and recursively separate two or more stochastic processes whose spectra overlap. We apply the Kalman filter to estimate phase data by optimally separating atmospheric phase noise and the phase information from a source. However, the performance of the Kalman filter critically depends on the accuracy of the model for each process under consideration. Therefore, mathematical or statistical models of the atmospheric phase data and source phase data are needed. Combining the time series modeling method with other radio astronomy data processing algorithms such as calibration and CLEAN, stochastic difference models for both source structure and phase noise can be found. Then, the Kalman filter is used for the phase estimation. In this paper, the basic procedure for correcting phase error by modeling and Kalman filtering is introduced and some results of simulated tests which are close to practical applications are given. Some considerations, advice, suggestions for using this technique, and further developments are given at the end of this paper.

INTERFEROMETER AND ATMOSPHERIC DISTURBANCE

The fundamental idea behind an interferometer or a synthesis array is that it measures the Fourier transform of the observed brightness distribution by cross-correlating the signals from antennas separated by distances up to thousands of kilometers. The measured complex visibility (unnormalized) is:

$$V(u,v) = \int T_B(x,y) e^{-i2\pi(ux+vy)} dx dy \quad (9.1)$$

where the $T_B(x,y)$ is the radio source brightness temperature distribution. The antenna beam pattern was omitted from (9.1) for simplicity.

The resolution of an interferometer increases as the baseline increases. However, the resolution of earth-based interferometers is ultimately limited by atmospheric turbulence. The phase measured by an interferometer is the phase difference related to the group delay between the arrival of the signal wavefront at each of the two antennas. In the noise free case, the group delay varies with the distance between two antennas and the position of the radio source relative to the baseline. But in practice, the measured delay is also affected by several other factors such as the receiving-system noise, bandwidth, and atmospheric turbulence. According to Fourier transform theory, a phase error $\Delta\phi$ in the frequency domain will cause a position shift in the spatial domain. For a simple case at one instant of time

$$V(u,v) e^{-i2\pi(\Delta\phi u + \Delta\phi v)} \longleftrightarrow T_B(x-\Delta\phi, y-\Delta\phi). \quad (9.2)$$

Moreover, the phase fluctuation is in a random fashion so that the shift is random and unpredictable. For synthesized data, the random phase error will cause scattering of the source power. If the phase error is on the order of the phase changes in the source visibility, we will have severely distorted maps. It is well known that the phase error at centimeter wavelengths is dominated by atmospheric water vapor which affects the refraction index of the traversed medium. Consider the refractive index $n(\mathbf{r},t)$ in a nonionized atmosphere:

$$n(\mathbf{r},t) = n_0(\mathbf{r}) + N_f(\mathbf{r},t) \quad (9.3)$$

where $n_0(\mathbf{r})$ is a time average of $n(\mathbf{r},t)$ at location \mathbf{r} and $N_f(\mathbf{r},t)$ represents the random fluctuation in $n(\mathbf{r},t)$. If two antennas of an interferometer are located at \mathbf{r}_1 and \mathbf{r}_2 , respectively, the phase error due to a nonhomogeneous medium is

$$\Delta\phi = k \int n(\mathbf{r}_1,t) d\mathbf{r}_1 - k \int n(\mathbf{r}_2,t) d\mathbf{r}_2 \quad (9.4)$$

where $k=2\pi/\lambda$, λ is a wavelength, and the integration range is the atmospheric path through which the wave travels. The interesting term from (9.4) is

$$\Delta\phi_f = k \int N_f(\mathbf{r}_1,t) d\mathbf{r}_1 - k \int N_f(\mathbf{r}_2,t) d\mathbf{r}_2 \quad (9.5)$$

which causes the random behavior in phase. With some assumptions and approximations, Han [1980] showed that the mean square value of $\Delta\phi_f$ has the following behavior

$$E[(\Delta\phi_f)^2] \text{ (troposphere)} \propto (D/\lambda)^2 \quad (9.6)$$

where D is the baseline length. To obtain a high resolution map, D/λ should be increased. However, as D/λ increases, the phase error due to water vapor in troposphere will increase and degrade the quality of the map. Therefore, a correction for phase error is necessary.

MODELING ALGORITHM

A simple and practical auto-regressive (AR) modeling approach is introduced for finding models of random processes. There are many text books available for the details of time series modeling theory [Box and Jenkins, 1976; Pankratz, 1983; Fuller, 1976; Abraham and Ledolter, 1983; Graupe, 1984]. Previously, time series modeling theory was applied to VLA phase data by Basart and Zheng [1986].

Assume that a class of noise-free phases received by an interferometer can be characterized by a real-valued finite random field $\mathbf{S} = \{S(t); 0 < t < T\}$ where T is the observation time length. A random atmospheric noise field $\mathbf{V} = \{V(t); 0 < t < T\}$ and a random system noise field $\mathbf{N} = \{N(t); 0 < t < T\}$ are introduced in the observation field so that

$$Z(t) = S(t) + V(t) + N(t) \quad (9.7)$$

where $\mathbf{Z} = \{Z(t), 0 < t < T\}$ is an observed random field. Thus a specific observed phase value is a realization of \mathbf{Z} value at time t :

$$z(t) = s(t) + v(t) + n(t). \quad (9.8)$$

Let the deterministic factors such as phase differences due to the geometric delay $g(t)$ and the position $p(t)$ of a source center be removed from (9.8). We have

$$y(t) = x(t) + v(t) + n(t) \quad (9.9)$$

where $y(t) = z(t) - g(t) - p(t)$ and $x(t) = s(t) - g(t) - p(t)$. The large scale variation of phase existing in (9.8) is removed in (9.9). The removal is necessary to obtain a working data sequence which is close to wide-sense stationary for a reasonable time duration. Generally, $v(t)$ is colored with a nonzero mean and $n(t)$ is white with a zero mean. We assume that the mean of $v(t)$ is removed by the empirical calibration. $v(t)$ is a main error factor to be removed.

Given noisy data $y(t)$, we want to estimate $x(t)$ in (9.9). To do this, models describing each process in (9.9) are required. Presently, no useful closed computational form has been found for each process in (9.9). Since the atmospheric phase variation is correlated in time, it is natural to think about using stochastic prediction time series models to describe the phase fluctuation. A family of stochastic models such as ARIMA (autoregressive integrated moving average) model was introduced to describe a variety of random phase behavior [Basart and Zheng, 1986]. In our previous work, a complete model identification and parameter estimation procedure was presented.

Conceptually, a specific ARIMA model will represent a random process with a specific covariance sequence. Given a covariance sequence of a random process, an ARIMA model can be identified and estimated for the process. In this work, because of the computation time, we shall primarily use AR models instead of ARIMA models for simplicity. We feel that this simplification is reasonable for testing purposes. One always can improve this technique by a more complete

modeling procedure. In the following, we shall introduce several parameter estimation methods with certain given conditions.

Parameter Estimation with Unknown Signal and Knowledge of Noise Covariances

Since atmospheric phase fluctuation is correlated in time, without losing generality, $v(t)$ can be described by an AR(q) process

$$v(t) = \sum_{i=1}^q \theta_i v(t-i) + a(t) \quad (9.10)$$

where $a(t)$ is an iid($0, \sigma_a^2$). (9.10) is a stochastic difference equation whose unknown parameters can be found from the given covariance.

Taking the covariance operation of both sides of (9.10), we have

$$C_v(h) = \sum_{i=1}^q \theta_i C_v(|i-h|) + \delta(h) \sigma_a^2 \quad (9.11)$$

where $C_v(h)$ is the autocovariance of $v(t)$ at lag h . Letting h vary from 1 to q , a set of q linear equations (Yule-Walker equation) will be formed for q unknown parameters of θ_i . A recursive algorithm can be used to solve the linear equations (Levinson, 1947; Jayant and Noll, 1984). Letting $h=0$, the variance of $a(t)$ can be obtained after θ_i is solved. Two useful formulas relating the covariances to parameters of an AR model are

$$C_v(0) = \sigma_n^2 / (1 - \theta_1^2) \quad \text{if } q=1 \quad (9.12)$$

$$C_v(0) = (1-\theta_2)\sigma_n^2 / (1+\theta_2)[(1-\theta_2)^2 - \theta_1^2]. \quad \text{if } q=2 \quad (9.13)$$

To find a model for $x(t)$, we take the covariance operation of (9.9) to get

$$C_y(h) = C_x(h) + 2C_{xv}(h) + C_v(h) + \delta(h)\sigma_n^2 \quad (9.14)$$

where σ_n^2 has usually been specified for a specific receiver. The cross terms between $n(t)$ with other processes are zero because $n(t)$ has a distribution of $iid(0, \sigma_n^2)$. $C_{xv}(h)$ is a cross covariance term for $x(t)$ and $v(t)$. $C_y(h)$ is a autocovariance of $y(t)$ which can be directly calculated from the observed data. Given $C_v(h)$, the autocovariance $C_x(h)$ of $x(t)$ can then be estimated from (9.14) if the $x(t)$ and $v(t)$ are uncorrelated with each other.

With little loss of the generality, $x(t)$ can be described by an $AR(p)$ process

$$x(t) = \sum_{i=1}^p \phi_i x(t-i) + w(t) \quad (9.15)$$

where $w(t)$ is $iid(0, \sigma_w^2)$ and the parameters, ϕ_i , can be estimated by solving the Yule-Walker equation. σ_w^2 can be estimated in the same way to obtain σ_a^2 . One should be careful that σ_w^2 , σ_a^2 , and the covariance matrix of $x(t)$ are semi-positive definite. Define a state variable vector $\mathbf{x}_t = \{x_j(t)\}; j=1,2,\dots,(p+q)\}$ and let $x(t) = x_1(t)$ and $v(t) = x_{p+1}(t)$. We have

$$\begin{bmatrix} x_1(t) \\ x_2(t) \\ \vdots \\ x_p(t) \\ x_{p+1}(t) \\ x_{p+1}(t) \\ \vdots \\ x_{p+q}(t) \end{bmatrix} = \begin{bmatrix} \phi_1 & & & & & & & \\ \phi_2 & & & & & & & \\ \vdots & & & & & & & \\ \vdots & & & & & & & \\ \phi_p & 0 & \dots & 0 & & & & \\ \hline & & & & 0(q,p) & & & \\ & & & & \theta_1 & & & \\ & & & & \theta_2 & & & \\ & & & & \vdots & I(q-1) & & \\ & & & & \vdots & & & \\ & & & & \theta_q & 0 & \dots & 0 \end{bmatrix} \begin{bmatrix} x_1(t-1) \\ x_2(t-1) \\ \vdots \\ x_p(t-1) \\ x_{p+1}(t-1) \\ x_{p+2}(t-1) \\ \vdots \\ x_{p+q}(t-1) \end{bmatrix} + \begin{bmatrix} w(t) \\ 0 \\ \vdots \\ 0 \\ a(t) \\ 0 \\ \vdots \\ 0 \end{bmatrix} \quad (9.16)$$

Equation (9.9) becomes

$$y(t) = x_1(t) + x_{p+1}(t) + n(t). \quad (9.17)$$

The forms of (9.16) and (9.17) are required for the Kalman filter. AR models, or ARMA models, can also be written in more general forms such as power spectral forms [Zheng, 1985] for other conventional digital signal processing techniques.

Parameter Estimation with Unknown Noise and Knowledge of Signal Covariances

If there is a way to obtain some statistical information on the phase variation caused by source structure, the unknown parameters in (9.10) and (9.15) can be estimated by the same way as the first method except for exchanging the roles of signal and noise. The required statistical information about signal is the covariance of the phase variation of the source.

Parameter Estimation with Unknown Signal and Knowledge of Noise
Variance

The condition of knowing only the noise variance is closer to practice than the condition of knowing all the covariances. In this case, if the atmosphere varies rapidly and the integration time is long, the phase error will be less correlated in time [Zheng, 1985]. Therefore, the phase noise has a very high portion of whiteness. In this case, the noise reduction on the white part is still helpful for reducing the phase noise. In practice, the variance of atmospheric phase fluctuation may be either given by the measured data of a calibrator source or by an empirical rms-baseline formula considering weather conditions. In this situation, the parameter estimation procedure is the same as the first method except that all $\theta_i=0$ and $q=0$ in equations (9.10) to (9.17).

PARAMETER ESTIMATION COMBINING THE TIME SERIES MODELING METHOD WITH
OTHER DATA PROCESSING ALGORITHMS

The above three methods giving general and basic modeling schemes require different a prior knowledge. This a prior knowledge is necessary for model identification and Kalman filtering; otherwise, we can do nothing for the phase with this technique. However, the required knowledge is statistical information which we may obtain in a variety ways. Using information given by other radio astronomy data processing algorithms, we may obtain such required information. Two methods listed below are examples of such approaches.

Parameter Estimation with Data from Calibrator Sources

The variance or autocovariance of the phase noise is required by methods A and C in the last section. This variance or autocovariance may be approximately estimated from phase data of calibrator sources in an empirical calibration process. During the empirical calibration process, data of a strong point-like calibrator source located near the unknown source observed are periodically collected. Since the noise free phase data of a point source is completely decided by the geometric delay and the shift distance of the source position from the phase center, any other fluctuations in the phase are due to noise. Therefore, the data from a calibrator source are used as correction factors applied to unknown source data in an empirical calibration procedure. The correction factors are not perfect because the

calibrator source is observed at a different position and at a different time than the observed-known source and the atmospheric condition varies in both spatial and time domains. However, the variance (possibly autocovariance) of the phase noise can be estimated from data on the calibrator source which provides statistical information about the atmosphere. Conceptually, even though the phase data of the calibration source are far from perfect to use for correction factors because of the variation in space and time, the variation of statistical information such as the variance can remain in a reasonable range for a reasonable time duration and certain spatial distance. Moreover, according to the nature of the empirical calibration, the estimated variance will be smaller than what it should be in most cases. In this case, the estimated variance of the phase noise will be less than what it should be, and then, the Kalman filter removes part of the noise and the results are reliable. In an off-line situation, with variances or covariances of phase error from data of the calibrator source, we always can use the methods mentioned in the last section to correct the data.

Parameter Estimation with CLEAN Components

According to the nature of the CLEAN deconvolution method, the Fourier transform of the first set of positive CLEAN components will provide the phase information of the source position and the large structure of an extended source. Let the phase of the CLEAN components be denoted by $c(t)$, the noise free phase of the unknown source is

$$s(t) = c(t) + e(t) \quad (9.18)$$

where $e(t)$ is an error term between $c(t)$ and $s(t)$ and it is a term to be estimated. The observation equation becomes

$$y(t) = c(t) + e(t) + v(t) + n(t). \quad (9.19)$$

The $g(t)$ had been presumably removed in (9.19). Moving $c(t)$ to the left side, we have

$$r(t) = y(t) - c(t) = e(t) + v(t) + n(t) \quad (9.20)$$

where $r(t)$ can be readily obtained which is the difference between the observed data and the phase of CLEAN components. $r(t)$ is also called a residual sequence with some information $e(t)$ existing inside.

Therefore, there is a need to extract the information from $e(t)$ instead of considering it being useless. Given the autocovariances of $v(t)$ and the variance of $n(t)$, AR models of both $v(t)$ and $e(t)$ can be estimated. Accordingly, the Kalman filter is applied to separate $e(t)$ and $v(t)$ from $r(t)$. Consequently, the estimation of $s(t)$ can be obtained with (9.18). The autocovariance of $v(t)$ may be obtained from various possible methods such as from the data of the calibrator source which mentioned above.

PHASE ESTIMATION BY KALMAN FILTER

The Kalman filter grew out of Wiener filter theory. The Wiener filter was developed to solve optimal estimation problems such as separating random processes whose spectra overlap. Classical frequency filters cannot solve this problem. A Wiener filter is basically a weighting function based on the Minimum-Mean-Squared-Error (MMSE) criterion. However, the Wiener filter solution does not lend itself very well to the discrete-data problem [Brown, 1983]. In 1960, R. E. Kalman provided a state-space method for formulating the MMSE filter [Kalman, 1960; Kalman and Bucy, 1961]. The two main features of the Kalman filter that are additional to the Wiener filter are 1) vector modeling of the random processes under consideration so that the multiprocess problem can be solved, and 2) recursive processing of the measurement data so that it can be used in real time and require little memory.

If the parameters of the stochastic difference equations of random processes in equations (9.10) or (9.20) are known, it is straight forward to apply the Kalman filter to separate these processes. Rewriting (9.15) as

$$\mathbf{x}_t = \Phi_t \mathbf{x}_{t-1} + \mathbf{w}_t, \quad (9.21)$$

the observation equation is

$$y(t) = \mathbf{H}_t \mathbf{x}_t + n(t) \quad (9.22)$$

where \mathbf{x}_t is a $((p+q) \times 1)$ process state vector at time t , Φ_t is a $((p+q) \times (p+q))$ state transition matrix which is given in (9.16) (the coefficient matrix), \mathbf{w}_t is a $((p+q) \times 1)$ white driving vector, and $\mathbf{H}(t) = \{h_i; i=1,2,\dots,p+q\}$ in which $h_i=0$ except for $h_1=1$ and $h_{p+1}(t)=1$. The covariance matrix of \mathbf{w}_t is given by $\mathbf{Q}(t) = \{Q_{ij}(t); 1 \leq i \leq p+q, 1 \leq j \leq p+q\}$ in which $Q_{ij}(t)=0$ except for $Q_{11}(t)=\sigma_w^2$ and $Q_{(p+1)(p+1)}=\sigma_a^2$. $n(t)$ is the system white noise with a variance of σ_n^2 . If $v(t)$ is colored, the measurement variance $R=\sigma_n^2$. If $v(t)$ is white, only the first p states are needed in equations (9.16) or (9.21), and then, $R = \text{variance of } v(t) + \sigma_n^2$.

Let the estimated state vector of \mathbf{x}_t be $\mathbf{x}_{t|t}$, the Kalman filter estimates $\mathbf{x}_{t|t}$ by minimizing the major diagonal of the error covariance matrix

$$\mathbf{P}_{t|t} = E[(\mathbf{x}_t - \mathbf{x}_{t|t})(\mathbf{x}_t - \mathbf{x}_{t|t})^T] \quad (9.23)$$

where the elements along the major diagonal are the error variances of the states. As the solution of the minimization, the Kalman filter equations can be derived. They are:

1. Predicting the state vector and the error covariance matrix

$$\mathbf{x}_{t|t-1} = \Phi \mathbf{x}_{t-1|t-1} \quad (9.24)$$

$$\mathbf{P}_{t|t-1} = \Phi \mathbf{P}_{t-1|t-1} \Phi^T + \mathbf{Q} \quad (9.25)$$

2. Computing the Kalman gain vector

$$\mathbf{K}_{t|t} = \mathbf{P}_{t|t-1} \mathbf{H}^T [\mathbf{H} \mathbf{P}_{t|t-1} \mathbf{H}^T + R]^{-1} \quad (9.26)$$

3. Estimating the state vector

$$\mathbf{x}_{t|t} = \mathbf{x}_{t|t-1} + \mathbf{K}_{t|t} [y(t) - \mathbf{H} \mathbf{x}_{t|t-1}]. \quad (9.27)$$

4. Evaluating the error covariance matrix

$$\mathbf{P}_{t|t} = [\mathbf{I} - \mathbf{K}_{t|t} \mathbf{H}] \mathbf{P}_{t|t-1}. \quad (9.28)$$

The principal idea is given at step 3 where the estimate is obtained by a sum of a prediction and a correction factor. The correction factor is given by the modified residual where the modification factor depends on the Kalman gain. The error variance of estimation is given by $\mathbf{P}_{t|t}$. With a sequence of $y(t)$ and all estimated parameters, given the initial values of $\mathbf{x}_{1|0}$ and $\mathbf{P}_{1|0}$, an estimated sequence $\mathbf{x}_{t|t}$ can be obtained by iteratively processing the above equations from step 1 to step 4 from the first sample of $y(t)$ to the last one. In our case, the estimated phase data are given by the first element of the vector $\mathbf{x}_{t|t}$. The phase closure concept may be used in phase estimation and only the independent phase sequences are filtered.

SOME EXPERIMENTAL RESULTS OF T-CALIBRATION

The applicability of the T-calibration was tested with some simulations for correcting phase data. The data used are from VLA and Hat Creek Millimeter interferometer (here after Hat Creek array).

A Simulation of Filtering Antenna Gains

Antenna gains provide correction factors for observed data. One of first tests was to estimate accurate antenna gains. If the Kalman filter can provide accurate antenna gains, it is equivalent to giving good estimates of the phase. Moreover, the computational load of estimating antenna-based gains is less than that of estimating baseline-based visibility data. Also, the visibility data are expected to be unchanged. With this in mind, we did a simulation to correct the antenna gains by the T-calibration method.

Figure 9.1 is a VLA map of the point source 1741-038. It was observed with a 6-cm wavelength on May 21, 1982 by R. A. Sramek for an atmospheric phase disturbance study. The rms values of phase fluctuation of twenty-four time series varied from 3° to 23° . The data were averaged over one-minute intervals for 207 minutes. The data were processed by the VLA conventional data processing programs. The peak flux density in Figure 9.1 is 0.4239 Jy/beam. Figure 9.1 is acknowledged as an original map for our simulation purpose.

The noise map was obtained by adding white noise to the gain file of sixteen series. The noise variance is 28.6° and the rms ratio of the original data and noise vary from 0.105 to 0.803. By completing the same data processing as for the original data, the map of the noisy data is obtained and is shown by Figure 9.2 where the peak flux density dropped to 0.4085 Jy/beam. This occurred because of the energy scattered due to phase error. With the noise variance such as 28.6° , the VLA data processing program did tremendously reduce the noise. However, the difference between the two maps can still be shown by the very low contour levels where the background noise has the same order as the low level structure of the original map.

The modeling and Kalman filtering procedure in the case of known noise variance was used to estimate antenna gains. Given variances of a noise $v(t)$ and system noise variance, the covariances of noise free phase sequence $x(t)$ were estimated from (9.14). Accordingly, the AR(p) models of antenna gains described by equation (9.15), where the p is as high as 3, were obtained by solving the Yule Walker equation. Then, a bank of Kalman filters of (9.24) to (9.28) was applied to estimate gains with the system equations of (9.21) and (9.22). Each antenna gain sequence was estimated by an individual Kalman filter with a specific set of parameters. The rms of the noise remaining in the antenna gains range from 3° to 15° . The related rms ratios between the original data and the noise remaining were increased by a factors from 3.6 to 6.2. By the same data processing programs used for the last two

maps, the map obtained after filtering the antenna gains is given by Figure 9.3. The peak flux was restored as 0.4234 which is very close to the original one and the low contour structure of Figure 9.1 and Figure 9.3 are remarkably similar.

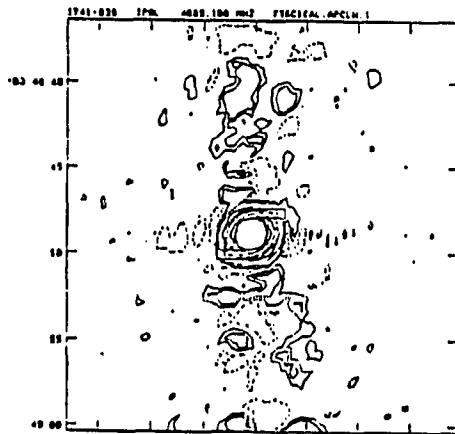


Figure 9.3. The filtered map from Figure 9.2. The peak flux density is 0.4234 Jy/beam. The contour levels are the same as Figure 9.1

Correcting Baseline-Based Phase Data

The T-calibration was tested for correcting the baseline-based phase data. The source map was chosen among the maps available to us. The source was a point source 3C48. The data were collected by R. A. Sramek for an atmospheric phase disturbance study in April 7, 1984 with the VLA in the C array and 2 cm wavelength. The observing atmospheric conditions were bad. The data record length was about 120 minutes and

the highest peak-to-peak fluctuation was more than 250° . Two thirds of the 351 correlator outputs had an noise rms above 24° and the worst case was 43° . Though the phase of this data set varied rapidly in time, the mean of the data sequence remained close to stationary. The original map obtained by the conventional data processing programs in VLA is shown in Figure 9.4 which has a dynamic range of 362. In this particular case, the atmospheric varies so rapidly that the phase noise has a high component of whiteness and the phase noise reduction can be significantly done by only removing the white component of the noise. The variance of the white component can be estimated from the ARMA analysis of the phase data. Consequently, the covariance of the non-white component could be found. Therefore, the parameters of model (9.15) were estimated and a bank of Kalman filters was applied to correct the phase data. Then, the same data processing procedure for the original data was used for the filtered data. The map of the filtered data is given by Figure 9.5. The dynamic range was increased to 657 and the edges of the source is sharper than those in the original map.

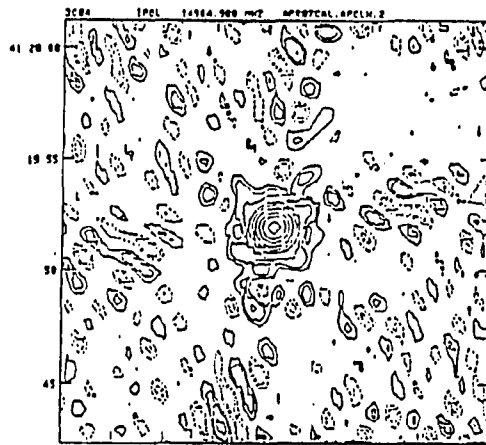


Figure 9.4. The observed map of radio source 3c84. The peak flux density is 0.5602 Jy/beam. The contour levels are: -1.0, -0.5, -0.3, 0.3, 0.5, 1.0, 4.0, 8.0, 16.0, 32.0, 50.0, 80.0, 100.0

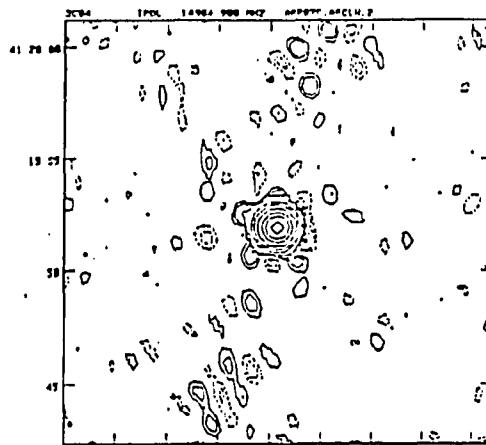


Figure 9.5. The filtered map from Figure 9.4. The peak flux density is 0.6022 Jy/beam. The contour levels in % of the peak value are the same as in Figure 9.4

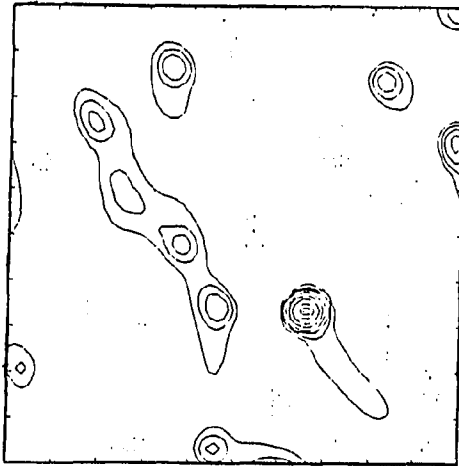
A Simulation of Filtering Phase Data Combining with the CLEAN

T-calibration was tested by a simulation for parameter estimation using CLEAN components and noise covariances. The data were collected for a radio source by three antennas in the Hat Creek array. The observation time was ten hours with three antennas in an observing frequency of 91.71 GHz (3.27 mm). The integration time was 100 second. The data were calibrated before we used them.

Figure 9.6 shows a CLEAN map made by the multi-channel mapping programs of Berkeley's radio astronomy laboratory. For simulation purposes, Figure 9.6 was acknowledged as an original map and any structure on the map was acknowledged as the original structure. A noisy map was obtained by adding noise on the original phase data and a restored map was obtained by reducing the phase noise. We want the restored map to be as close as possible to the original map.

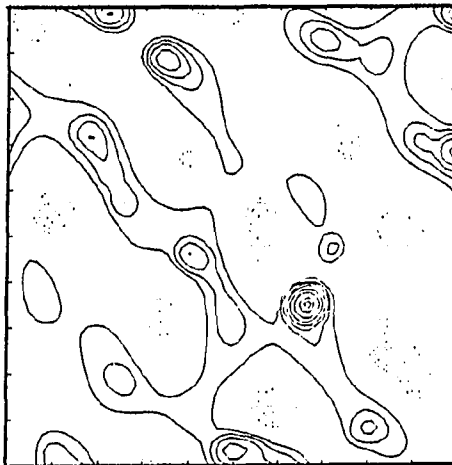
The range of intensities of Figure 9.6 is from -0.012 Jy/beam to 0.112 Jy/beam and the ratio of peak to rms of background noise (hereafter DR) is about 116.5. In the frequency domain, there are three visibility data sequences because of three antennas used. The noise map was obtained by adding three computer generated AR(1) noise sequences to the phase data. The variances of noise range from 11° to 18° . The noisy map is shown by Figure 9.7 where the range of intensities is from -0.021 Jy/beam to 0.099 Jy/beam and the DR is about 77.8. The peak intensity dropped to 0.099 Jy/beam from 0.112 Jy/beam because of the phase noise. A procedure including parameter

estimations for phase models and phase correction in time order was used to restore the noisy map.



01.09: Original map -20-16-8-4-2 1 2 4 8 16 32 64 85 99

Figure 9.6. The original map. The peak flux density is 112.5 mJy. The contour levels in % of the peak value are: -20, -16, -8, -4, -2, 1, 2, 4, 8, 16, 32, 64, 85, 99



01.12: Noisy map -20-16-8-4-2 1 2 4 8 16 32 64 85 99

Figure 9.7. The phase noise added map from Figure 9.6. The peak flux density is 93.5 mJy. The contour levels in % of the peak value are the same as in the Figure 9.6

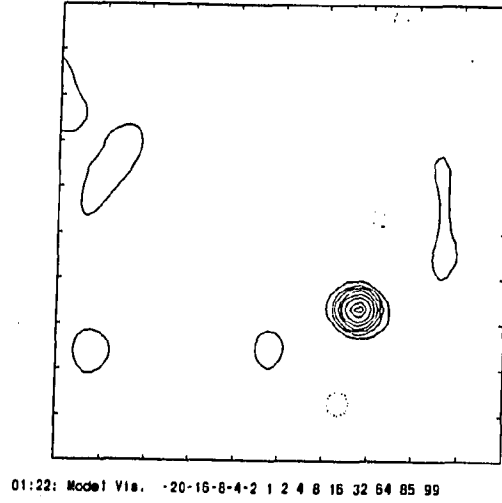


Figure 9.8. The map made with the first fifteen CLEAN components. The peak flux density is 127.2 mJy. The contour levels in % of the peak value are the same as in the Figure 9.6

This method uses the information provided by the CLEAN method and the empirical calibration method. The first fifteen CLEAN components of the noisy map (Figure 9.7) are positive. A map of them, shown in the Figure 9.8, has an intensity range from -0.004 Jy/beam to 0.127 Jy/beam with a DR is of 164. Since we specified that Figure 9.6 was the original map, we see that both Figure 9.7 and Figure 9.8 are severely distorted. The first fifteen positive CLEAN components of the noise map were used to find $c(t)$ of equation (9.18). Then, equation (9.20) was formed and its covariances were estimated from $r(t)$. Given covariances of noise, the covariance of $s(t)$ could be readily found. AR(1) models were used to model $e(t)$,

$$\hat{e}(t) = \phi_1 \hat{e}(t-1) + k u(t) \quad (9.29)$$

where $\hat{e}(t)$ is an estimate of $e(t)$, and $u(t)$ is an iid(0,1). ϕ_1 was given by

$$\phi_1 = c_e(1) / c_e(0) \quad (9.30)$$

and K was found by

$$k^2 = c_e(0) / (1 - \phi_1^2). \quad (9.31)$$

Consequently, AR(1) models for both $e(t)$ and $v(t)$ were estimated by equations (9.30) and (9.31). For antenna pair 1-2,

$$\hat{v}(t) = 0.80\hat{v}(t-1) + 0.31u(t) \quad (9.32)$$

$$\hat{e}(t) = 0.17\hat{e}(t-1) + 0.66u(t) \quad (9.33).$$

For antenna pair 2-3,

$$\hat{v}(t) = 0.81\hat{v}(t-1) + 0.33u(t) \quad (9.34)$$

$$\hat{e}(t) = 0.24\hat{e}(t-1) + 0.43u(t) \quad (9.35)$$

For antenna pair 3-1,

$$\hat{v}(t) = 0.69\hat{v}(t-1) + 0.32u(t) \quad (9.36)$$

$$\hat{e}(t) = 0.30\hat{e}(t-1) + 0.43u(t) \quad (9.37)$$

where $\hat{v}(t)$ is an estimate of $v(t)$. Each pair of $v(t)$ and $e(t)$ was written in the state space form of (9.16), and then, three Kalman filters given by (9.24) to (9.28) were applied to the noisy phase data to estimate $\hat{e}(t)$ s. Then the estimated phase data were obtained by

$$\hat{s}(t) = \hat{c}(t) + \hat{e}(t) \quad (9.38)$$

The map restored by the filter is given by Figure 9.9.

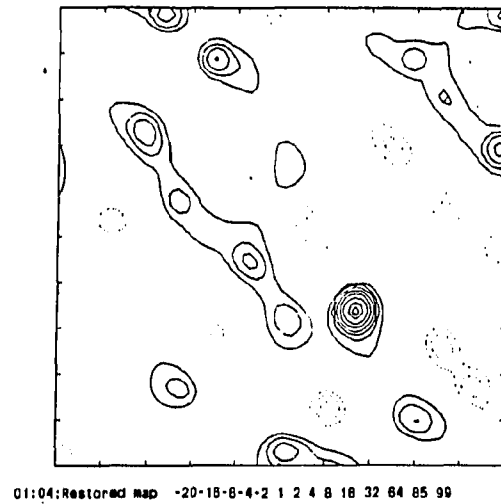


Figure 9.9. The restored map from Figure 9.7. The peak flux density is 115.4 mJy. The contour levels in % of the peak value are the same as in the Figure 9.6

The noise variances actually used were smaller than the measured values of $v(t)$ by an amount of 50%. The range of intensities in the restored map are from -0.012 Jy/beam to 0.115 Jy/beam, and the DR is about 125. All these statistical values of the restored map are closer to the original one than the noisy one. Since we specified Figure 9.6 as an original map, the restored map given by Figure 9.9 was expected to be as close as possible to the Figure 9.6. Comparing the noisy map (Figure 9.7), the CLEAN component map (Figure 9.8), and the restored map (Figure 9.9) to the original map (Figure 9.6), the restored one is closest to the original map. Again, the algorithm worked well.

SUMMARY

A new technique has been developed for correcting the phase error caused by atmospheric disturbances for long-baseline interferometer arrays. The simulations were done for correcting antenna-based gains and for correcting baseline-based phase data.

The atmospheric induced phase fluctuation was found to be autocorrelated in time. This provided the fundamental grounds for modeling and filtering the phase data in time. This technique used time series AR modeling theory to obtain the phase models for both noise and noise free phase processes. Time series information from other radio astronomy data processing methods, such as calibration and CLEAN, was used for the model's parameter estimation. For simplicity, AR(p) models, where $p \leq 3$, were used for both the signal and noise models. A simple but practical parameter estimation procedure, such as solving the Yule-Walker equation, was used to find parameters of AR models.

Given AR models of both noisy and noise free phase processes, the Kalman filter is was employed to optimally separate these random processes. The Kalman filter was used because it could separate the random processes whose spectra overlapped in a situation where only part of the statistic information was available. Moreover, the Kalman filter provided a measure of the goodness of the estimation which is not available in other radio astronomy phase data correcting algorithms.

The weakness of this T-calibration is that some a priori statistical information about the random processes under consideration should be known or measured.

The a priori knowledge required of this technique are autocovariances of the processes of interest. They can be partly provided by empirical calibration and CLEAN components. The CLEAN components provide global structure information of a source but they introduce non-ignorable phase error also. Hence, using the CLEAN components as a signal model in the calibration is not mathematically perfect so a correction term should be added to the signal model. Therefore, we proposed a mathematical model which combined the model of the CLEAN components and a correction term estimated from the residual sequence of the phase difference between observed data and data from CLEAN components. In this model, no approximation was made. The problem is, can we obtain the correct term from the residual sequence? At this moment, in our T-calibration technique, time series modeling and Kalman filtering is used to estimate the correction term. This estimation is an approximation. However, the sum of the data from the CLEAN components and from the estimated correction term provides a better signal model than that from the CLEAN components alone.

T-calibration should be further developed and tested. There are a few places where this technique needs to be improved. A more careful and automatic variance estimation algorithm should be developed so that the estimation procedure always provides stable and reasonable

solutions. Another problem is that a more complete and robotic model identification procedure needed to be developed. In this future modeling procedure, parsimonious models with proper ARMA forms and orders can be estimated for each random process under consideration.

Another future development of T-calibration is to develop a new calibration scheme that combines the T-calibration and self-calibration so that the data estimated from the T-calibration provides source models for self-calibration. This scheme may be necessary to properly maintain the closure relation which may be violated by time series filtering, or is not optimally maintained during the filtering.

Other future work of interest is to develop a phase estimation scheme where incorporating microwave water-vapor radiometer data into the Kalman filter for estimating and correcting atmospheric phase disturbance.

ACKNOWLEDGMENTS

This project was partially funded by National Science Foundation Grant no. AST-8217135. We thank R. A. Sramek, T. J. Cornwell, and R. D. Ekers of VLA for helpful discussion throughout this project. We thank Dr. Sramek for sharing VLA phase data. We thank C. H. Wright and W. J. Welch for providing Hat Creek interferometer data and multi-channel mapping programs. We thank Wright for providing help for installing and modifying the multi-channel mapping programs. We thank the National Radio Astronomy Observatory for providing initial financial support for the project and for providing continued service support.

REFERENCES

- Abraham, B., and J. Ledolter. 1983. Statistical Methods for Forecasting. John Wiley & Sons, New York.
- Armstrong, J. W., and R. A. Sramek. 1982. Observations of tropospheric phase scintillations at 5 GHz on vertical paths. Radio Science 17:1579-1586.
- Basart, J. P., and Y. Zheng. 1986. Modeling very large array phase data by the Box-Jenkins method. Radio Science 21:863-881.
- Baars, J. W. M. 1967. Meteorological influences on radio interferometer phase fluctuations. IEEE Trans. Antennas Propagation AP-15:582-584.
- Basart, J. P., G. K. Miley, and B. G. Clark. 1970. Phase measurements with an interferometer baseline of 11.3 km. IEEE Trans. Antennas and Propagation AP-18(3):375-379.
- Box, G. E. P., and G. M. Jenkins. 1976. Time series analysis: forecasting and control. 2nd ed. Holden-Day, San Francisco, CA.
- Brown, R. G. 1983. Introduction to random signal analysis and Kalman filtering. John Wiley & Sons, New York.
- Clark, B. G. 1980. An efficient implementation of the algorithm "CLEAN". Astron. Astrophys. 89:377-378.
- Cornwell, T. J. 1983. A method of stabilizing the clean algorithm. Astron. Astrophys. 121:281-285.
- Dravskikh, A. F., and A. M. Finkelstein. 1979. Tropospheric limitations in phase and frequency coordinate measurements in astronomy. Astrophys. and Space Sci. 60:251-265.
- Fuller, W. A. 1976. Introduction to Statistical Time Series. John Wiley and Sons, New York.
- Graupe, D. 1984. Time series analysis, identification and adaptive filtering. Robert E. Krieger Publishing Company, Malabar, Florida.
- Hamaker, J. P. 1978. Atmospheric delay fluctuations with scale sizes greater than one kilometer, observed with a radio interferometer array. Radio Sci. 13:873-891.
- Han, W. J. 1980. Effects of atmosphere irregularities on radio interferometer measurements. Chinese Astronomy 4:174-184.

- Hargrave, P. J., and L. J. Shaw. 1978. Large-scale tropospheric irregularities and their effect on radio astronomical seeing. *Mon. Notic. Royal Astron. Soc.* 182:233-239.
- Hinder, R. A. 1970. Observations of atmospheric turbulence with a radio telescope at 5 GHz. *Nature* 225:229-253.
- Hinder, R., and M. Ryle. 1971. Atmospheric limitations to the angular resolution of aperture synthesis radio telescopes. *Mon. Notic. Royal Astron. Soc.* 154:229-253.
- Hogbom, J. A. 1974. Aperture synthesis with a non-regular distribution of interferometer baselines. *Astron. Astrophys. Suppl.* 15:417-426.
- Hogbom, J. A. 1984. CLEAN as a pattern recognition procedure. Pages 247-254 in J. A. Roberts, ed. *Indirect Imaging*. Cambridge University Press, Cambridge, NY.
- Ishimaru, A. 1978. Wave propagation and scattering in random media. Vol. 2. Academic, Orlando, Fla.
- Jayant, N. S., and P. Noll. 1984. Digital coding of waveforms: principles and applications to speech and video. Prentice-Hall, Englewood Cliffs, New Jersey.
- Jennison, R. C. 1958. A phase sensitive interferometer technique for the measurement of the Fourier transforms of spatial brightness distributions of small angular extent. *Mon. Notic. Royal. Astron. Soc.* 118:276-284.
- Kalman, R. E. 1960. A new approach to linear filter and prediction problems. *Trans. ASME* 82:35-45.
- Kalman, R. E., and R. S. Bucy. 1961. New results in linear filtering and prediction. *J. Math. Phys.* 25:261-278.
- Levinson, N. 1947. The Wiener rms (root mean square) error criterion in filter design and prediction. *J. Math. Phys.* 25:261-278.
- Mathur, N. C., M. D. Grossi, and M. R. Pearlman. 1970. Atmospheric effects in very long baseline interferometry. *Radio Sci.* 5:1253-1261.
- Moran, J. M., and B. R. Rosen. 1981. Estimation of the propagation delay through the troposphere from microwave radiometer data. *Radio Sci.* 16:235-244.
- Pankratz, A. 1983. Forecasting with univariate Box-Jenkins models. John Wiley, New York, NY.

- Readhead, A. C. S., and P. N. Wilkinson. 1978. The mapping of compact radio sources from VLBI data. *Astrophys. J.* 223:25-36.
- Resch, G. M., D. E. Hogg, and P. J. Napier. 1984. Radiometric correction of atmospheric path length fluctuations in interferometric experiments. *Radio Sci.* 19:411-422.
- Schwab, F. R. 1980. Adaptive calibration of radio interferometer data. *Proc. Soc. Photogr. Optic. Instrum. Eng.* 231:18-25.
- Schwarz, U. J. 1978. Mathematical-statistical description of the iterative beam removing technique (Method CLEAN). *Astron. Astrophys.* 65:345-356.
- Tatarskii, V. I. 1961. Wave propagation in a turbulent medium. Part I. McGraw-Hill, New York, NY.
- Tatarskii, V. I. 1971. The effects of the turbulent atmosphere on wave propagation. U.S. Dept. of Commerce, National Technical Information Service, Springfield, Va.
- Treuhaft, R. N., and G. E. Lanyi. 1987. The effect of the dynamic wet troposphere on radio interferometer measurements. *Radio Sci.* 2: 251-265.
- Wesseling, K. H., J. P. Basart, and J. L. Nance. 1974. Simultaneous interferometer phase and water vapor measurements. *Radio Sci.* 9:349-353.
- Zheng, Y. 1985. Modeling VLA phase data and improving the synthesized image with univariate Box-Jenkins models and Kalman filters. M.S. Thesis. Iowa State University, Ames, Iowa.

GENERAL SUMMARY

Various techniques have been developed and applied to data and image processing in radio astronomy and nondestructive evaluation. The subjects of these techniques are: statistical analysis, ARMA modeling, stochastic filtering, deconvolution, and feature extraction. The general objective of developing and applying these techniques is to improve the quality of data and images of radio astronomy and nondestructive evaluation.

A model based filtering scheme has been developed to enhance the industrial NDE X-ray radiographs and radio astronomy radio source maps (Part III, Part V, and Part VIII). This scheme includes AR modeling and Kalman filtering processes. Knowledge about noise of an image can be obtained from either the image forming mechanism or a flat region of the image. Therefore, the covariances of both signal and noise can be estimated with the given system equation and noisy measurements. Consequently, AR descriptions about signal and noise can be obtained and the Kalman filter can be used to separate the signal and noise processes. This filter scheme will not destroy the details of an image and the background contents as other smoothing filters will.

A fast and effective maximum entropy deconvolution method has been developed to reconstruct industrial NDE X-ray and IR images (Part VII). The MEM image is obtained by maximizing the entropy of a reconstructed image constrained by the χ^2 statistic. A Newton-Raphson iteration method is used to find an extremum of the entropy subject to the

condition enforced by Lagrange multipliers. This MEM has been shown that it gives smooth NDE images with shape edges.

Some other image enhancement techniques have been applied to improve NDE images (Part II). They are: the sigma filter, median filter, adaptive smooth filter, trend removal, and histogram equalization.

Some image feature extraction methods have been developed to detect features of NDE images (Part II, Part IV). They are: the modified masking function, thresholding, and low-level image segmentation. These methods have been used to detect flaws of industrial NDE X-ray radiographs.

A rule-based expert system was discussed to automate the image segmentation, modeling, and Kalman filter processes (Part IV). A set of knowledge rules and a set of control rules were used to monitor and control the entire image enhancement procedure without human interaction. It is useful for industrial applications.

A new technique has been developed for phase data correction of a synthetic aperture antenna array or an interferometer (Part I and Part IX). Atmospheric-induced phase disturbance has been found to be correlated in time which can be used to obtain ARMA descriptions of the phase data. Given a proper system model which describes the observations, consisting of signal (radio source information) and noise (atmospheric-induced phase error), we can find an ARMA signal model and an ARMA noise model, respectively. This is done by first estimating

covariances of the signal process and covariances of the noise process using the given observations. The Kalman filter, incorporating the CLEAN algorithm, is used to correct the phase data. Mathematically, this method is more advanced than the self-calibration method since the CLEAN components are not perfect for the signal model. Furthermore, since this method corrects data in time, it can apply to an interferometer while self-calibration can not do so. By combining this method with self-calibration, a new calibration method can be readily obtained.

ACKNOWLEDGMENTS

I wish to express my deepest appreciation to my mentor and friend, Professor John Philip Basart, who has provided valuable guidance, enthusiastic encouragement, and steady support for me throughout my graduate education in the last five years, and for the preparation of this dissertation. Working with him, I was very fortunate to have many wonderful opportunities to be involved in various research projects and to gain valuable experience.

I wish also to express my deep appreciation to the director of the NSF University/Industrial center, Dr. Donald O. Thompson, who has provided many suggestions and support for my research projects in NDE.

I thank the National Science Foundation (NSF⁵), the NSF University/Industry Center⁶ for NDE at Iowa State University, and the Electrical Engineering and Computer Engineering Department of Iowa State University for funding the various projects in which I was involved.

I am greatly indebted to my friend, Dr. John F. Aurand for his consistent encouragement and spiritual support in most of the time of my graduate studies.

⁵NSF Grant No. AST-8217135.

⁶The center is sponsored by the NSF and various aerospace, energy, transportation and equipment manufacturing industrial companies and it is operated by the Ames Laboratory, an Energy & Mineral Resources Institute of the Department of Energy.

I am also greatly indebted to the chairman of our department, Dr. J. O. Kopplin, for providing an excellent and inspiring academic research environment. I am very grateful to our signal and image processing group members for their help and discussion in the preparation of this manuscript. Those are, in alphabetical order: Patrick Anderson, Ed Doering, Eric Dunn, Charly Gatot, Yong S. Koh, Brian Lovewell, Dick Wallingford, and Chen-Yee Wong. Especially, I would like to thank Chen-Yee Wong and Yong S. Koh. Without their help, I would not be able to have the dissertation be typed on time.

I am grateful to the director of the VLA, Dr. R. D. Ekers for providing partial travel fees, the opportunity of using the VLA instrument, and many suggestions. I would also like to express my appreciation to the VLA staff for many valuable technical suggestions, sharing VLA data, and using VLA facilities. They are: Dr. R. A. Sramek, Dr. T. Cornwell, A. Braun, R. Kummerer, Dr. J. Von. Gorkam, R. Payne, D. Retallack.

I am indebted to the Director of Radio Astronomy Laboratory of University of California at Berkeley, Dr. W. J. Welch and Dr. C. H. Wright for providing the Hak Creek Millimeter interferometer data and the Multi-Channel Mapping Programs.

I thank James Lloyd and David Fischer of Martin Marietta, Daniel King of McDonnell Douglas Aircraft Company, and Dave Utrata of the Association of American Railroads for providing industrial IR images and X-ray radiographs.

I am grateful to the members of my committee, Dr. Steve F. Russell, Dr. Donald O. Thompson, Dr. R. G. Brown, and Dr. Wayne A. Fuller for their valuable time. Especially, I would like to express my appreciation to Dr. Brown and Dr. Fuller for their signal processing and statistical time series analysis courses. I thank Dr. A. A. Read for providing help in the use of his AGRAPH program. I also thank Dr. W. Q. Meeker for providing help in the use of his TSERIES program. I thank Dr. J. Gray for the valuable technique discussions about NDE radiographic images.

It is with the great pleasure that I thank my uncles, Wing L. Lee and Henry Feng. Their support and encouragement made my study in the United States possible. I would like to extend this appreciation to my uncles and aunts: Chang-Xuan Zheng, Yee Feng, Chang-Lian Zheng. I thank Mr. Wen-Yuan Wang for many educational talks and encouragement. Many thanks to my brother, Gang Xu, who always has confidence in me: "Brother, you are always able to walk through the way you choose."

Special thanks are due to Xiao-Bin Zhong, for the spiritual support in an important period of my graduate study.

I wish to take this opportunity to express my deepest appreciation to my elementary school and middle school teachers in China, Zhong-Yu Wang, Jin-Xia Qiu, and Shi-Chun Shen, who did their best to enlighten and edify me during the most difficult time period in China in 1960s and 1970s.

Last but not least, my heartfelt and endless thanks to my parents, who always stand behind me and who consistently encourage and support me. Thanks to my natural father, Qi Feng, and my natural mother, Chang-Bi Zheng, who were unfairly persecuted for more than twenty years since I was one year old, who made sacrifices in order to let me have a secure growing environment, who have suffered the endless pain of family separation because of the political persecution. I wish to express my deepest appreciation to my foster father, Ren Xu, and my foster mother, Chang-Qing Zheng. I owe a debt of thanks to my foster parents for the love they have given to me, for the education they have provided for me, for the standard of excellence they have instilled in me, and the faith they have always had in me. Without the support of my foster parents, my study in the United States would have been impossible.